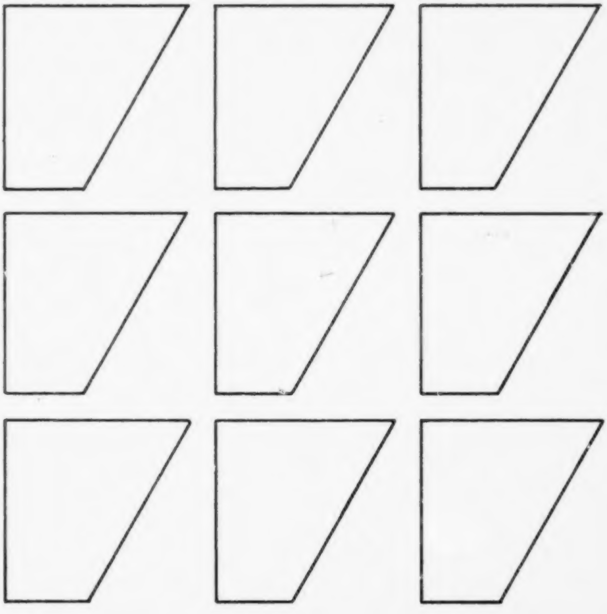
6/11/85 U3515-2.6-F

Monograph Series in Finance and Economics

Monograph 1984-1/2

The U.S. Payments System: Costs, Pricing, Competition and Risk

By David Burras Humphrey



Salomon Brothers Center
for the Study of
Financial Institutions

Graduate School of
Business Administration
New York University

SPONSORING SUBSCRIBERS TO THE MONOGRAPH SERIES IN FINANCE AND ECONOMICS

Bankers Trust Company
Bernstein-Macaulay, Inc.
The Chase Manhattan Bank, N.A.
Chemical Bank
Citibank, N.A.
Cowen & Company
The Dreyfus Corporation
The Equitable Life Assurance Society of the United States
Exxon Corporation
The First National Bank of Chicago
The Guardian Life Insurance Company of America
Manufacturers Hanover Trust Company
Mellon Bank
Metropolitan Life Insurance Company
Morgan Guaranty Trust Company
Phibro-Salomon, Inc.
The Philadelphia National Bank
The Prudential Insurance Co. of America
Salomon Brothers, Inc.

THE MONOGRAPH SERIES IN FINANCE AND ECONOMICS

Editor: Anthony Saunders

The Monograph Series of the Schools of Business of New York University specializes in the publication of original studies in the fields of finance and economics. It provides an outlet for manuscripts which are too long and too broad in scope for the scholarly journals or of too great length and depth for the business periodicals. The submission of manuscripts in these fields is solicited.

For subscription rates and single copy orders, see inside back cover of Monograph.

The Monograph Series is the successor to the *Bulletin* (1925-1964), which was generously supported by C.J. Devine as one of the activities of the Institute of Finance. The Monograph Series is published by the Salomon Brothers Center for the Study of Financial Institutions at the Graduate School of Business Administration of New York University. Editorial offices are located in 1307 Merrill Hall, 90 Trinity Place.

SALOMON BROTHERS CENTER FOR THE STUDY
OF FINANCIAL INSTITUTIONS
Director: Arnold W. Sametz

NEW YORK UNIVERSITY GRADUATE SCHOOL
OF BUSINESS ADMINISTRATION
Dean: Richard R. West

Biographical Note

David B. Humphrey is Chief, Financial Studies Section, Research and Statistics Division, Board of Governors, Federal Reserve System. Prior to entering the Federal Reserve in 1975, he taught economics at San Francisco State and Tulane Universities. He has published numerous articles in professional journals, is an Associate Editor of the *Journal of Monetary Economics*, and holds a Ph.D. degree from the University of California (Berkeley). Over the last five years, he has been involved in economic analyses of Federal Reserve payment services, concentrating on statistical estimation of scale economies in check, ACH, and wire transfer processing operations. His current work has focused on economic cost/benefit studies of new payment services, such as electronic check collection, a direct settlement service, and a universal return item service.

Abstract

The U.S. Payments System: Costs, Pricing, Competition and Risk

This study is an economic analysis of the U.S. payments system. The analysis concentrates on the impact that the Depository Institutions Deregulation and Monetary Control Act of 1980 (MCA) has had on the costs and pricing of payments instruments. The main topics covered include:

Costs:

- Estimates of the user and social costs of cash, checks, ACH (automated clearing house), and wire transfer of funds payment methods;
- Estimates of the costs (to banks) and the revenues (to the U.S. Treasury) of the MCA;

Pricing:

- Analysis of methods used to price Federal Reserve payment services and float:
- Analysis of the so-called Private Sector Adjustment Factor (PSAF) or "markup" to Federal Reserve payment service prices;
- Estimates of scale economies in Federal Reserve (wholesale) processing of cash, checks, ACH, wire transfer, and book-entry U.S. government security payment operations;
- Estimates of scale economies in bank (retail) payment related operations in collecting and purchasing deposits and in making and disbursing loans;

Competition:

- Estimates of Federal Reserve and private sector market shares in processing check, ACH, and wire transfer payment instruments;
- Analysis of nonprice aspects of payments services, and potential new payment arrangements, which can affect competitive structure in the future;

Risk:

- Analysis of current developments regarding large dollar wire transfer networks as a result of efforts to reduce risk of a settlement failure; and
- Analysis of the likely market response to final adoption of bilateral net credit limits, sender net debt caps, payments finality, or assessment of fees for the provision of daylight credit on large dollar payment networks.

All of these topics, except for the estimation of Federal Reserve and commercial bank scale economies, have not been heretofore publicly addressed or only addressed in a very limited way.

Monograph 1984-1/2

The U.S. Payments System: Costs, Pricing, Competition and Risk

By David Burras Humphrey

TABLE OF CONTENTS

I. Introduction 1

II. A Description of the U.S. Payments System 3

- **A. The Evolution of Checks, Wire Transfers, and Settlement** 3
- **B. Current Use of Checks, ACH, Wire Transfers, and Settlement** 5
  - Use of Different Payment Methods 5
  - Float 7
  - Payments Systems in Other Countries 8
- **C. Risk in the Payments System** 10
- **D. The Payments Mechanism and the Conduct of Monetary Policy** 10

III. Costs and Prices of Payments Services 13

A. User Versus Social Costs of Different Payments Instruments 13

B. Commercial Bank Prices for Payments Services 19

C. Direct Fees Versus Balance Requirements 21

IV. The Pricing of Federal Reserve Payments Services 24

A. The Quid Pro Quo of Federal Reserve Pricing 24

The Federal Reserve Membership "Problem", Deposit Coverage Ratio, and Monetary Control 24

Increased Treasury Revenues from the Monetary Control Act 26

B. Federal Reserve Pricing and the DIDC: Two Forces for Change in Payments Practices 26

The Cost-Push and Demand-Pull Effects of Federal Reserve Pricing 26

Federal Reserve Prices for Payment Services 30

The DIDC and the Pricing of Payments Services 32

C. Choice of a Pricing Strategy for Federal Reserve Payment Services 33

Marginal Cost Pricing 35

Optimal Departures from Marginal Cost Pricing 36

Federal Reserve Market Sensitive Pricing 39

D. Other Pricing Strategies 40

Peak-Load Pricing 40

Par Value Pricing 42

Benefit-Flow Pricing 43

Two-Part Pricing 44

E. Similarities Between Correspondent Bank and Federal Reserve Pricing Strategies 45

V. The Private Sector "Markup" and Pricing Federal Reserve Float 46

A. Pricing Federal Reserve Services with a Private Sector "Markup" 46

B. The Pricing of Federal Reserve Float 50

Reducing Float Versus Pricing Float 50

The Real Resource Cost of Pricing or Reducing a Transfer Payment 51

Pricing of Float 53

Monetary Policy Effects 55

VI. Scale Economies in Providing Payments Services 56

A. Wholesale Payments Operations: Scale Economies in Federal Reserve Payments Services 56

Use of Cross-Section Data 56

A Cost Function Model 57

B. Estimates of Federal Reserve Scale Economies 58

A Look at the Raw Data 58

Check Processing Scale Economies 59

ACH Scale Economies 63

Wire Transfer Scale Economies 66

Cash And Book-Entry Securities Processing Scale Economies 66

Scale Economies and Choice of a Pricing Strategy 67

C. Retail Payments Operations: Scale Economies in Providing Commercial Bank Deposit and Loan Services 68

Average Costs of Retail Payment and Intermediation Operations 68

Scale Economies in Commercial Banking 70

VII. Competition in Providing Wholesale Payment Services 74

A. Existing Competitive Structure 74

Check Market Share 74

ACH Market Share 75

Wire Transfer Market Share 77

B. The Importance of Quality and Other Product Characteristics 78

C. Potential New Payment Services 79

Check Truncation 79

Electronic Check Collection (ECC) 79

Direct Settlement Service (DSS) 80

Value Dating and Same-Day Settlement 81

VIII. Risk in Payments Systems 82

A. Daylight Overdrafts on Large-Dollar Funds Transfer Networks 82

FedWire Daylight Overdrafts 82

CHIPS Daylight Overdrafts 83

Other Large Dollar Networks 86

Reducing FedWire Daylight Overdrafts Through Reductions in Overnight Funding 86

B. Current Developments in Reducing Interbank Risk on Large-Dollar Payment Networks 89

Payments Examinations and Access to Net Settlement 90

FedWire: Finality of Payments and Risk Reduction 91

CashWire: Finality of Payments and Risk Reduction 93

CHIPS: Finality of Payments and Risk Reduction 94

CHESS: Finality of Payments and Risk Reduction 94

C. Future Developments in Reducing Interbank Risk 94
Likely Future Developments 94
Making Receiver Guarantees Acceptable on CHIPS 95
Three Principles of Cost/Risk Assignment 96
Sender Net Debit Caps 97
D. Other Alternatives for Reducing Interbank Risk (and Reasons for Their Disfavor) 99
Rolling Settlement 99
Settlement Insurance 100
Indemnification Agreements 101
Prearranged Collateralized Loan with a Commitment Fee 101
E. Examples of Risk Management Arrangements Outside of Banking 102

IX. The Impact of Costs, Competition, and Risk in Changing the Payments Mechanism 105
A. Costs: Paper *vs.* Electronics 105
ACH as a Substitute for Check Processing 106
B. Competition: National *vs.* Local Services 107
Retail Payment Services 107
Wholesale Payment Services 107
Future Growth in Transactions Volume 108
Effects of Interstate Banking 109
Reduced Use of Checks 110
C. Risk: Provisional *vs.* Final Settlement Systems 110
Future Structure of Large-Dollar Networks 111
Competitive Position Between Banks 111
Pricing Interbank Risk and Cost Recovery 112
Institutional Changes in Financial Markets 113

X. Summary and Conclusions 116
The Costs of Different Payments Instruments 117
The Effect on Bank Costs from the Monetary Control Act (MCA) 119
Scale Economies and the Pricing of Federal Reserve Payment Services 119
Clarifying the Private Sector Adjustment Factor (PSAF) 122
Market Shares and Competition in Providing Payment Services 124
Current Developments in Reducing the Risk of a Settlement Failure 126
Market Response to Risk Reduction on Large-Dollar Networks 128

Appendix: A Translog Cost Function Model for Estimating Federal Reserve Payments Processing Scale Economies 130

Bibliography 133

List of Tables

1. Volume, Value, Growth, and Average Value per Transaction by Payment Method 6
2. Breakdown of User and Social Cost of Payment Instruments by Cost Components 14
3. Commercial Bank Fees for Payment Services Provided to Respondent Banks 20
4. The Monetary Control Act and Treasury Revenues 27
5. Federal Reserve Prices for Payment Services 31
6. Federal Reserve Float: Level and Distribution 53

7. Federal Reserve Check, ACH, and Wire Transfer Unit Costs by Four Size-Classes of FR Offices 59
8. Federal Reserve Check Scale Economies Over 1976-1983 60
9. Scale Economies at Federal Reserve ACH Offices Over 1978-1982 65
10. Average Cost (per account and per dollar) of Five Banking Services 71
11. Volume and FR Market Share in Check, ACH, and Wire Transfer Wholesale Payments Services 75
12. Actual Funds Transfer Daylight Overdrafts on FedWire and CHIPS 85
13. Daylight Overdrafts as a Percent of Federal Funds Transactions and Reduction in Daylight Overdrafts if Multi-Day or Continuing Contracts were Used 88

List of Figures

1. New or Renewed Local Check Clearinghouses Activated Since FR Pricing of Payment Services in 1981 29
2. The Relationship Between Marginal Cost, Average Cost, and Price Under Different Scale Economy Situations 34
3. Average Cost Curves by Bank Size Class: The Production and Intermediation Approaches to Measuring Bank Output 73
4. Illustrative Reserve Account Balance by Time of Day for an Institution with a FedWire Daylight Overdraft 84
5. Schematic Representation of Payments Over FedWire, CashWire and CHIPS 92

Copyright 1984 by New York University

I. Introduction*

The passage of the Depository Institutions Deregulation and Monetary Control Act of 1980 (MCA) began a process which is fundamentally restructuring the U.S. payments system. The final users of check, automated clearing house (ACH), and large-dollar wire transfer services are currently and will be increasingly influenced by the changes underway, although most are largely unaware of the precise nature of the payments changes now occurring. Specifically, the deregulation of deposit interest rate ceilings combined with the pricing of Federal Reserve Bank payments services and related developments is having and will continue to have the following effects on the payments system:

- A movement away from providing consumers with "free" retail payments services toward charging direct fees for services used;
- Loss of Federal Reserve float benefits to check writers (a loss upwards to $6 billion dollars per day in interest-free investable funds);
- Increased incentives to use (potentially) more efficient electronic payments delivery systems;
- Reductions in the amount of time consumers must now wait until they can use funds deposited in their accounts (delayed availability); and
- Increased costs and reduced flexibility in making domestic (and possibly international) large-dollar funds transfers (averaging $2.5 million per transaction).

These payments system developments are far reaching in their impact, both in terms of redistributing the benefits and costs of payments system use and influencing the future direction of payments innovation. For the most part, these developments are due to the following policies;

- The MCA requirement that Federal Reserve check, ACH, wire transfer, and net settlement services be directly priced, along with the removal of interest ceilings on certain deposits;

---

*The views expressed are solely those of the author and do not in any sense imply concurrence by the Board of Governors or its staff. Discussions with the following individuals is gratefully acknowledged: Allen Berger, Bill Dudley, Ed Ettin, Carl Gambs, Morgan Hallmon, Elliott McEntee, Gene Snyder, Bruce Summers, and Florence Young. These individuals bear no responsibility for errors which may exist in this document. Research assistance was performed by Oscar Barnhardt and Bryan Davis with typing by Brenda Harley. Editing and encouragement to finish were provided by Stacie Humphrey and Ernie Bloch.

- The MCA directive to price and/or eliminate Federal Reserve float;
- The way in which Reserve Bank payment services have been priced, along with the apparent scale diseconomies in large check processing operations *vs*. scale economies in ACH (or electronic) funds transfers;
- The reconfiguration of current transportation networks for check presentment along with check truncation and return item pilot programs; and
- The adoption of risk reduction measures on large-dollar payments networks.

The impact that these causative policy factors are having or will have on the payments mechanism are in turn directly related to the four economic issues focused upon in this monograph:

(1) The underlying *costs* of providing different types of payments services, including the scale economies associated with their production;
(2) The *pricing* strategies developed to assign costs and other benefits to users of payments services;
(3) The structure of *competition* and market shares between provide~~rs of~~ (wholesale) payments services; and
(4) The nature of the alternatives being considered for reducing *risk* in large-dollar funds transfer networks.

These four issues, preceded by a brief description and overall view of the U.S. payments system, are discussed below in detail. Although some knowledge of payments operation is assumed, the discussion is kept largely at the lay level. Complex statistical/econometric analyses are summarized and presented in a nontechnical manner.

II. A Description of the U.S. Payments System

The U.S. payments system provides for the transfer of funds between economic agents. Almost all funds are transferred by one (or more) of the following means: (1) cash; (2) paper checks or electronic payment orders drawn upon demand deposit accounts at banks; and (3) wire transfers between banks using correspondent balances or Federal Reserve reserve accounts. The Federal Reserve System is actively involved in all three types of transfers. It distributres coin and currency used in cash transactions and maintains their fitness. Along with private clearinghouses and large banks, it processes and provides settlement for paper-check transactions. It is the main supplier of services for the domestic wire transfer of funds within the United States, and it provides net settlement facilities for privately-operated U.S. domestic and international wire transfer services.

Virtually all of these payment services have private sector counterparts. Armored carriers operate their own coin and currency sorting facilities and transport cash to and from major users. Correspondent banks and service bureaus process checks and perform related accounting services for their corporate and financial institution customers. Groups of large and medium-sized banks have established jointly owned facilities which send domestic and international large-dollar wire transfers. Domestic and international transfers are now usually settled through use of reserve accounts at Reserve Banks but some transfers are still settled through use of reciprocal holdings of bank correspondent balances.

A. The Evolution of Checks, Wire Transfers, and Settlement

Paper checks developed as a substitute for currency transactions and have played an important role in the U.S. payments system since 1865. Checks were cleared, and funds transferred, through a network of interbank correspondent balances. In order for one bank's checks to be honored when deposited at another institution, a correspondent balance was debited and credit given to the check depositor. Correspondent balances were established and maintained through the shipment of gold (and later currency) between banks. Shipping costs were basically proportional to the value of the currency or gold being shipped (and to the weight, in the case of gold). When reciprocal balances were insufficient, credit risk was incurred by the bank cashing a check. Out of this process evolved a system of check-exchange charges that were proportional to the par or face value of the check being deposited. When exchange charges were deducted, checks were paid at less than par value (non-par banking). Efforts to minimize exchange charges often led to circuitous routing of checks, which impeded the flow of funds and commerce.[1]

Payments and financial transactions, reflecting the concentration of real economic activity, were restricted largely to regional money centers. Financial transactions among these regional centers were subject to domestic exchange rates. The existence of domestic exchange rates, where it may have cost a payor in Chicago $1.05 to purchase a $1.00 item from a seller in New York, reflected the cost (and risk) of settling a funds transfer between money centers. No centralized settlement facility existed and gold or currency was the settlement medium used.

The establishment of the Federal Reserve System in 1913 altered the U.S. payments system in two important respects. First, Federal Reserve member banks could transfer funds by wire using a single, centralized reserve-account balance. They thus had less need to maintain a complex network of correspondent balances to clear checks. Although correspondent balances are still used today to clear checks of smaller banks, book-entry accounting using wire transfers between reserve accounts has all but eliminated the need to ship gold or currency between regional banks to settle check-expenditure flows.[2] Because of the existence of a centralized settlement facility, domestic exchange rates ultimately disappeared (although other transactions costs remained).

The second change in the payments system was the virtual elimination of non-par banking. This elimination was accomplished by a requirement, written into the Federal Reserve Act, that all checks cleared through the Federal Reserve be settled at par value. In sum, Federal Reserve participation in the payments mechanism increased payments efficiency (1) by providing a centralized mechanism for settling check transactions, and (2) by eliminating the incentive for circuitous routing of checks that had developed under non-par banking.

---

[1] Similar effects exist today through so-called remote disbursement. Payors gain float benefits by maximizing the time between when a check is issued or written and when it can be finally presented for collection. By accpeting remotely disbursed checks, payees in effect are discounting the face value of the check by the value of the interest on the lost float—a form of non-par banking.

[2] See, for example, Garbade and Silber (1979) for a more complete discussion of these developments.

B. Current Use of Checks, ACH, Wire Transfers, and Settlement

Use of Different Payment Methods. It is estimated that 40 billion paper checks, worth $36 trillion, were written on around 110 million checking accounts in 1983; the average value of a check transaction was $910 (see table 1).[3] The volume of wire transfers at the Federal Reserve and in the private sector represented less than 0.2 percent (57 million) of that of paper checks, but the average value per transaction was $2.5 million. Wire transfers supported a total expenditure flow of $142 trillion, four times that of checks. Currently, check volume is growing by 5 to 6 percent a year while wire transfer volume is rising by 11 percent annually.

The electronic transfer of funds using automated clearing houses (ACHs) has not yet reached mature volumes. It is viewed as a potentially lower-cost substitute for the system based on paper checks. ACH volume in 1983 was more than seven times that of wire transfers but still represented only 1 percent (400 million) of the 40 billion paper checks written. Growth has been rapid, however, averaging 27 percent a year over 1981-83.

There is an obvious hierarchy in the use of payments instruments, which is related to the average value per transaction for each payments method. As the average value per transaction rises, the method of payment moves from cash to checks to a wire transfer (see table 1).[4] Cash, composed of coin and currency, is used for very small dollar transactions since, for these purposes, the opportunity costs and risks are minimal. For transactions in the hundreds of dollars, however, the (opportunity) cost of holding idle cash balances rises substantially, as does the expected value of loss through theft. Although most demand balances by law receive no explicit interest payment and are in this sense like an idle cash balance to the account holder, check writers gain float while cash users give up earnings. In addition, the risk of loss in using checks is much less than in holding cash. Thus, as the value of the transaction rises, consumers substitute checks (ACH, NOWs, share drafts, and credit cards) for cash. And when very high transaction values are encountered—those in the hundreds of thousands or millions of dollars—a wire transfer offers the greatest safety due to the special security procedures involved (passwords, restricted entry facilities, dedicated communication lines, and sometimes encryption).

---

[3] In 1983, negotiable order of withdrawal accounts (NOWs), share drafts at credit unions, and demand balances at mutual savings banks accounted for approximately 18 percent of total check expenditures.

[4] ACH payments, NOW and Super NOW account drafts, share drafts, and credit card transactions are close substitutes for checks drawn on demand deposits. Available evidence on these payment methods indicates an average value per transaction not too different (although somewhat lower, except for ACH) than that for checks drawn upon demand deposits.

TABLE 1

Volume, Value, Growth, and Average Value per Transaction by Payment Method 1983

	Volume (millons)	Total Value ($ trillions)	Annual Growth (1981-1983)	Value per Transaction ($)
Cash[1]	112,000 (very approximate)	$ 2.8	9%	$25
Check[2]	40,000	36	6%	910
ACH[3]	400	.7	27%	1,800
Wire Transfer[4]	57	142	11%	2,500,000

[1] *Cash*. Cash transactions are not directly reported. An educated guess by a national commission put cash transactions at two-thirds of all types of transactions (National Commission on Electronic Fund Transfers, 1977, p. 333). For retail sales transactions, however, excluding food and drinking establishments, cash transactions are only 30% of total transactions (Curtin, 1983, p. 27). But this likely understates the true percentage as food stores and coin operated machines are not included. The estimate of the total value of cash transactions used in the table relies upon a new survey of currency and transaction amount usage (Survey Research Center, 1984) which indicated that the average family spends $192 per month in coin and currency transactions compared to $2,444 a month on check transactions (excluding checks used to obtain cash). Thus for every $12.73 spent via check, $1 is spent using cash ($192.00/$2,444 = $1/$12.73). Dividing the $36 trillion value for check expenditures in the table by $12.73 gives $2.8 trillion as an estimate of the total value of cash transactions. This, of course, assumes that the ratio in the survey ($192.00 in cash transactions to $2,444 in check transactions) holds for the nation as a whole. The average value per cash transaction in the retail store survey (Curtin, 1983) was $25, with $37 for checks. This check figure compares well with the median value of a check written in 1979 of $36 (Federal Reserve Bank of Atlanta, 1979, p. 178). Dividing the $2.8 trillion estimate of the total value of cash transactions by the $25 average value per transaction gives 112,000 million as an approximate estimate of the total number of cash transactions. This volume figure represents 73% of the total of all types of transactions shown in the table. The annual growth of cash transactions shown reflects the actual growth in coin and currency in circulation, assumed to equal the growth in transaction value.

[2] *Check*. The number of (private) checks written in 1979 was 32.8 billion and the growth rate between 1974-79 was estimated to be 5.6% annually (Federal Reserve Bank of Atlanta, 1979, p. 1). These data suggest that 40 billion checks would have been written in 1983. The average value per check transaction in the Atlanta study was $757 in 1979 (p. 73) and was increased to $910 for 1983 using the 20% inflation rate experienced between 1980 and 1983. The value per check transaction is skewed upwards due to a small number of exceptionally large dollar value checks which are written.

[3] *ACH*. ACH transaction volumes and growth are from the Federal Reserve's cost accounting publication (*Planning and Control System Expense Report*) called PACS, adjusted downward slightly to correct for transfers counted twice when handled by two Federal Reserve offices. The data includes private sector processing volume, which is almost entirely from the New York Clearinghouse. The value per commercial ACH transaction averaged $3,800. This number is strongly influenced by the ACH transfers which occur on the night cycle, which is primarily used for concentrating corporate cash balances from receiving banks to disbursing banks. The average ACH night cycle

In addition to the reduced risk of loss, another important distinction between payments instruments lies in the speed with which funds are available. Banks depositing checks with the Federal Reserve may obtain same-day credit or may be required to wait one to two days (at a maximum) before receiving credit, depending upon the location of the bank upon which the deposited check is drawn. More recently, while credit may be given immediately or in whole day increments, the depositing bank will have to pay for the float involved in collecting checks on distant banks. In contrast, a (Federal Reserve) FedWire wire transfer results in immediate availability of funds. Other wire transfers (CHIPS, BankWire, CashWire, CHESS, SWIFT—discussed below) also result in same-day funds availability.

A third distinction lies in credit reversibility or payments finality. A payment in cash, of course, is a payment in "good funds." But credits for deposited checks and certain ACH transfers can be reversed (insufficient funds at the bank upon which they are drawn account for 85 percent of all returned items). In contrast, (FedWire and CashWire) wire transfers are final or nonreversible when made. Thus, even if the bank sending the funds fails later in the day, the recipient is protected against loss.[5] In sum, wire transfers are a quicker and safer means for transferring funds than are paper checks. But wire transfers are more difficult to arrange and are over forty times more expensive (for the Federal Reserve) to process.

Float. Float exists because not all payment intruments involve the instantaneous (or same day) transfer of funds between payor and payee. Payments instruments can be classified into those which do generate float and those which do not. In the U.S., such a classification would be:

---

[5] Strictly speaking, immediate funds availability and nonreversibility are properties of the Federal Reserve's wire transfer service (Fedwire) and that of BankWire's CashWire service. They do not apply to wire transfers made over other networks, such as CHIPS. These important distinctions will be fully discussed below when payments mechanism risk is addressed.

---

transfer was $21,000, while day cycle ACH debits (credits) averaged $350 ($700) per transfer. The average government ACH transaction was $500. As government (commercial) ACH volume was 61% (39%) of total volume, the weighted average value per ACH transaction is $1,800 (i.e., .39 ($3,800) + .61 ($500) = $1,787). Multiplying $1,800 by 400 million transactions gives $.7 trillion as an estimate for total ACH value transacted.

[4] *Wire Transfer*. Wire transfer volumes are for FedWire and CHIPS only, as volume on other U.S. networks is small. Volume growth data are from FedWire (all FedWire data are found in PACS, adjusted downward to correct for transfers counted twice when handled by two Federal Reserve offices). Value per transaction reflects FedWire and CHIPS average values (CHIPS data are from unpublished sources). FedWire had 38 million transfers worth $85 trillion (or $335 billion per day over 255 business days). CHIPS had 19 million transfers worth $57 trillion (or, $225 billion per day over 255 days). The average value per transaction was $2.2 million for FedWire and $3.0 million for CHIPS.

Payment Instruments Where:

Float Exists:	Float Does Not Exist:
Checks (including NOW drafts, share drafts, money market mutual fund—MMMF—drafts)	Coin, Currency ACH Transfers Wire Transfers

Thus checks and check substitutes are the only payment methods which generate float.[6] Some float exists as well for financial instruments which require time to process prior to the collection and transfer of the funds involved, such as bond coupons, food stamps, matured savings bonds and U.S. government book-entry securities transfers. However, these instruments are excluded from the payments instruments discussed here.

Float is important because it drives a wedge between the private and social costs of a transaction. As such, it can strongly affect the relative competitiveness and distribution of market shares between the two main providers of check collection services—the Federal Reserve and large correspondent banks. That is, possibly lower Federal Reserve check collection prices can be offset by competing correspondent banks that offer speedier check collection. Here the benefit of reduced float (faster funds availability) for a depositor can offset some or all of the increased processing and transportation costs involved in such expedited collection. A second important aspect of float is its influence on the relative attractiveness of checks versus ACH or other electronic payments. Payors often go to great lengths to gain check float (by drawing their checks on banks located at considerable distances from payees—remote disbursement—so that collection times are lengthened) and only in certain instances have given up these benefits to make payments through an ACH (e.g., the direct deposit of payroll and the corporate trade payments programs, both discussed below).

Payments Systems in Other Countries. In all countries, cash transactions are believed to be the most numerous. After cash, consumers use checks for the bulk of their transactions in the U.S., Canada, the U.K., and Italy. In other European countries a different consumer payment system is dominant—the giro. A check is a debit transfer (where the payee, by depositing a check for collection, originates a debit to the payor's account) while the giro is a credit transfer (where the payor originates a credit to the payee's account). Checks take time to collect and so create float. However, payors using a giro do not gain float because the payment transaction involves both payment and collection at the same time. Giro transactions are like ACH credit transfers and comprise over fifty percent of all non-cash transactions for all of the European Economic Community members except Italy (21 percent) and the U.K. (19 percent). The U.S. has the lowest giro participation rate since less than one percent of its non-cash transactions are ACH credit transfers (Bank Administration Institute, 1980).

---

[6] This is a generalization, since some Federal Reserve processing float exists in the provision of coin and currency services to banks and in the ACH and wire transfer operations. This float value is small compared to check float and will be ignored until later.

Along with ACH credit transfers, one other U.S. "no float" payment method similar in operation to a giro would be a point of sale (POS) payments system operated by a debit card. On a real-time POS system, payment and collection also occur at the same time. POS systems are in very limited use today partly due to the loss of payor float, partly because they are expensive to set up, and partly because merchants favor shared access systems while most financial institutions favor limited-access proprietary systems. In the longer-run, POS systems are viewed as cost-effective substitutes for branching by financial institutions since the delivery of payment services and cash needs can be met more cheaply and conveniently through POS and automated teller machines (ATMS) than by building (or maintaining) branch office networks.

A final observation on the payments systems in different countries is the striking difference in direct payments system involvement by central banks. The U.S. is at one extreme with the Federal Reserve having the greatest degree of participation in regulatory, supervisory, and actual payments processing. This participation is a direct result of the low level of concentration which exists in the U.S. banking industry. In contrast, other nations have highly concentrated banking structures (e.g., the 6 largest Canadian banks hold 94% of total banking assets).

With a concentrated banking industry, what would in the present U.S. system be payments between many accounts at different financial institutions are internalized in banking systems of other nations and represent payments between accounts at different branches of the same institution. As a result banks, not the central bank, have been responsible for developing and providing check processing operations and wire transfers. Payments coordination between, say, ten banks is much easier than coordination between thousands, as would be the case in the U.S. with over 14,000 banks and 24,000 other types of depository institutions (savings and loan associations, mutual savings banks, credit unions). In the disaggregated U.S. banking system, the Federal Reserve (since 1913) and the various associations of correspondent banks (since the start of the New York Clearing House

Association in 1853) have coordinated payments between different financial institutions.

C. Risk in the Payments System

Risks exist in any payments system. There are risks due to fraud, operational breakdowns, accounting mistakes, and the unexpected failure of a payments mechanism participant to settle for funds transferred during the day. Of these four risks, the last one—settlement risk—has the greatest potential for precipitating a sequence of failures that, if it occurs, would severely disrupt the operation of financial and product markets unless emergency credit was provided through the Discoutnt Window at Federal Reserve Banks, who would then face a risk of loss (credit risk).

Settlement risk is more manageable when small dollar payment methods are used, such as cash, checks, ACH transfers, and credit cards. This is a direct result of the fact that small dollar losses are easier to absorb out of earnings flows or equity capital writedowns than are exceptionally large dollar losses. An important additional element, however, has to do with the certainty with which liability for losses are apportioned among interested parties for large dollar electronic payments.[7] In both areas—the size of the potential losses and the *ex ante* assignment of liabilities—the wire transfer payment method contains the greatest risk.

The risk issue, as it applies to large-dollar funds transfer networks, is discussed in detail below. This discussion outlines the possible ways in which this risk could be reduced and notes the impact the various risk-reducing measures could have on the operation of the payments mechanism, financial markets, and the future development of electronic payment networks.

D. The Payments Mechanism and the Conduct of Monetary Policy

The ultimate goal of monetary policy is to control or influence the level of real goods and services produced in an economy (aggegate demand). This end, of course, has effects on the level of employment, the rate of inflation, the balance of payments, productivity, and economic growth which were identified as being among the more important of the nation's economic goals in the Full Employment and Balanced Growth Act of 1978 (The "Humphrey-Hawkins Act"). The means used to achieve this end have been dichotomized in the popular press into money supply targeting or interest rate targeting. Intermediate targets such as these are used rather than targeting on levels or growth of the underlying real economic variables because, among other reasons, data on the former are much more timely and accurate than information on the latter. Under either the interest rate or the money supply (intermediate) targeting procedure, developments in the payments system are important; but they are more important for money supply targeting.[8]

---

[7] This legal issue was noted in a recent study for the Association of Reserve City Bankers (1982) p. 44 and appendix E.

Under money supply targeting, new payments instruments can create money supply measurement problems and/or affect velocity relationships and therefore the historical association between a money supply measure being targeted and the level of real GNP which occurs. New payments instruments causing the greatest difficulty are those which are designed to be close substitutes for checks. These check substitutes typically require lower idle balances or have balances which receive interest payments, thus blurring the theoretically convenient distinction between transactions and precautionary balances and their separate relationship with real output and employment. In both of these situations, the velocity relationships which underlie accurate money supply targeting is altered.

These velocity relationships could also be affected by shifts between checks and electronic funds transfer methods, such as ACH and/or wire transfers. The shift from checks to ACH has in fact occurred but the effect on velocity has been small. The value of ACH transactions is currently estimated to be only 2% of total check transactions. This estimate can be used to approximate the proportion of demand deposit type balances which may be turned over more rapidly due to ACH transfers, which can be structured to require smaller idle transaction balances than check payments.

More real transactions could be supported on a smaller demand deposit base if there were a shift from checks to wire transfers. Although price and float incentives do not now favor such a shift, such incentives could be dramatically altered if the Federal Reserve or the private sector adopts some form of electronic check collection for large dollar checks.

In summary, the payments mechanism effects generally asserted to have had the greatest effect on velocity, and therefore money supply targeting, are:

- New payments instruments (NOW accounts, super NOWs, share drafts, MMMF drafts); and

---

[8] With interest rate targeting, payments system developments—new payments instruments and payment techniques—are more easily dealt with than with money supply targeting. This is because undesired interest rate variations can be easily offset by daily Federal Reserve open market operations while the associated money supply measurement problems are more difficult to counteract accurately on a timely basis.

- New payment techniques (cash management, remote disbursement).

Both effects can be circumvented by adopting the so-called Divisia Index (weighted) measures of monetary aggregates rather than continue to rely on the existing simple sum measures published today. The Divisia Index procedure, while very complex, is in theory basically unaffected by the introduction of new payments instruments and the consequent substitution which then occurs between the new and existing instruments. It also can accommodate substitution between types of payments instruments brought about through use of new payment techniques, such as cash management. With cash management, checks are collected faster, idle demand balances are concentrated and reduced, and money market instruments are purchased using wire transfers in order to earn an interest return on what were previously idle balances. Thus money market instruments have been substituted for demand balances. One disappointing feature of the Divisia Index, however, is that there is only a marginal improvement in using it in traditional money demand analyses compared to the simple sum measures of money supply. Thus the benefit of capturing the effects of new payments instruments and new payment techniques through the Divisia Index may not offset its increased complexity in measurement, although work continues on reducing this problem. In any event, both the simple sum and the Divisia Index are computed and used internally at the Federal Reserve in reaching judgements concerning monetary policy.

III. Costs and Prices of Payments Services

A. User Versus Social Costs of Different Payments Instruments

The costs of the U.S. payments system can be viewed two different ways:

- User costs—the real resource and transfer payments borne by the individual or corporation which decides to use one payments method over another. User costs thus should include float costs/benefits and the opportunity costs of holding nonearning funds or balances; and
- Social costs—the real resource costs borne by society as a whole, which nets out any transfer payments (like float and idle cash holdings) between payments participants.

User costs are preferred if one wishes to investigate a major determinant of the demand for particular types of payment instruments and the inducements for substitution between payment methods. In this instance, the float benefits of using checks are important since there is no float gain in using cash or initiating an ACH or wire transfer. Check float thus lowers the user cost of checks. In contrast, the opportunity costs of holding and using nonearning cash assets as an "inventory" for cash transactions is also a transfer payment (to the government) but this time it is a user cost, not a benefit like check float.

While check float and cash opportunity costs can influence the decision to demand and use one type of payment method compared to another, both of these costs, viewed from society as a whole, are transfer payments. Check float represents a transfer payment from payees to payors while users of cash are transferring earning assets to the U.S. Treasury.[9] User costs net of transfer payments, but including all real resource costs involved in making the transfer, yields the true social or real resource costs involved in providing different payments services.

A breakdown of estimated payments costs by production, processing, float, and opportunity costs is shown in table 2. Three general conclusions can be drawn from looking at these cost compo-

---

[9] To illustrate, each time a consumer writes a check for cash, demand balances are transferred to the U.S. Treasury (by way of a bank's prior purchase of idle vault cash). In using these Treasury balances for government expenditures, one of which is the cost of minting coin and printing currency, less government borrowing is needed to sustain a given expenditure level. By providing coin and currency, the Treasury is in effect obtaining an interest free loan from cash users whose only real resource costs are the costs of minting, printing, and distribution. The opportunity cost of cash holdings less the real resource costs represents a net transfer payment.

TABLE 2

Breakdown of User and Social Cost of Payment Instruments by Cost Components
(1983: $ millions)

	Cash	Checks	ACH	Wire Transfer
Production Costs: (Costs to produce the instrument prior to use)	$327[1]	$ 1,250[4]	0	0
Processing Costs:				
a. Payor (Accounting/Mailing)	lost cash, theft	13,431[5]	$72[9]	$239[11]
b. Payee (Processing/Accounting)	7,044[2]	8,805[2]	21[9]	
c. Bank (Processing/Transportation		3,200[6]		
d. Federal Reserve (Processing/Transportation)	151	392[7]	44	58
Float Costs:	0	33,000[8]	6[10]	5[10]
Opportunity Costs of Holding Idle Funds:	13,604[3]	0	0	0
Total User Cost:	21,126	−5,922	131	302
Total Social Cost:	7,522	27,078	137	297
Unit User Cost:[12]	19¢	−15¢	33¢	$5.30
Unit Social Cost:	7¢	68¢	34¢	$5.21

[1] Production costs were estimated by multiplying the number of currency notes outstanding (9,684 million) times an average production cost of $23 per 1,000 notes produced at the Bureau of Engraving and Printing (giving $223 million). The costs of coin issue are around $.02 per coin for approximately 97.5 billion coins outstanding (giving $1,951 million). This production cost for all currency and coin outstanding is transformed into a yearly cost as follows. Since the average $1 bill is replaced every 1.5 years, with other denominations replaced at more infrequent intervals, all currency was assumed to be replaced every 2 years. Coin has a virtually indefinite lifetime but requires replacement over time as individual coins are lost. For our purposes, coin was assumed to be replaced every 100 years. To these replacement costs are added the cost of producing new currency and coin due to yearly growth in demand, a function of inflation and population growth. The yearly growth of cash has been 9%. In sum, the yearly production cost of cash is $327 million ($223 million/2 + $1,951 million/100 + ($223 million + $1,951 million) × .09 = $327 million a year).

[2] Payee costs of accepting cash and checks for retail sales includes point of sale and accounting costs, the costs of theft and loss of cash, safekeeping and security costs, along with deposit costs and fees paid to financial institutions (Curtin, 1983). These costs, expressed as a percentage of the average cash and check retail transaction amount, were 2% for cash and 2.5% for check. This percent cost per average transaction value also applies to the total share of retail sales paid for by cash or check, both of which shares were estimated to be 30% of total (non-food or drink) retail sales. Applying these 30% shares and the 2-2.5% cost percentages to all retail sales in 1983 of $1,174 billion yields a total payee cost of cash of $7,044 million (.02 × .30 × $1,174) and $8,805 million for check (.025 × .30 × $1,174). These payee cost estimates cover bank cash costs as well. Bank check costs, however, are only partly recouped by fees charged to retailers, since consumers and corporations are also directly served.

[3] The opportunity cost of holding idle coin and currency was derived from evaluating the $158 billion in coin and currency in circulation, which is less than the $172 billion outstanding, in 1983 by the average 90 day Treasury bill rate in 1983 (8.61%) from table A24 in the Federal Reserve *Bulletin*. This figure excludes coin and currency held by the U.S. Treasury and Reserve Banks but includes idle cash balances at depository institutions. The fact that vault cash at banks can be used to satisfy reserve requirements reflects the fact that effective reserve requirements were lowered in 1959, not the possibility that the seigniorage benefits to the government were reduced and that vault cash costs are now part of reserve requirements.

[4] Check production costs were estimated by taking the actual production costs per standard consumer type check (2.5¢ per item) and business type check (5.0¢ per item) for a large east coast check printing firm and multiplying these average production costs by the volume of the two types of checks written. Since roughly 75% of all checks produced in 1983 were consumer type checks and 25% were business type checks (including form checks for automatic dispersal), these ratios were multiplied by the 40 billion checks written in 1983 to estimate the volume of the two basic types of checks written. Thus consumer checks were estimated to cost $750 million and business checks $500 million, for a total check production cost of $1,250 million.

[5] Payor costs were composed of accounting and disbursing costs of business and government payors and postage ($.20) and envelope costs ($.02) for all payors (business, government, and consumer). The opportunity cost of consumer payors in taking time to write and mail checks was *not* included because few consumer payors actually have the opportunity of getting paid for the time saved if they did not write checks. Accounting and disbursing costs for business payors is estimated to be $.24 per check or payment transaction, based on the $.239 per payment transaction cost for the U.S. Treasury's direct deposit program (Dudley, 1984). This excludes all postage and Federal Reserve expenses. According to U.S. Postal Service sources, there were 119,381 million pieces of mail handled in fiscal 1983. Earlier analyses by the University of Michigan's Survey Research Center (1978 and 1980) indicated that 82.5% of all mail sent originated in the nonhousehold sector while 17.5% originated from households. Of nonhousehold origi-

nated mail, 35% was bill payment related and typically included checks sent to pay for bills received. In contrast, 75% of all household originated mail was bill payment related. Overall, some 42% ($.75 \times .175 + .35 \times .825 = .42$) of all mail (119,381 million items) is estimated to be payment related and, by assumption, check related. At \$.22 each, the 50,140 million in payment related items generates \$11,031 million in consumer, small business, and corporate payor mail costs. Since almost all of these mailed items are likely to be first class mail, and the number of first class items was 64,247 million in 1983, bill payment related mail comprises close to 78% of all first class mail in this estimate. Accounting and disbursing costs for business and government payors is estimated by multiplying the estimated per payment cost (\$.24) by the 25% of total checks written (.25 × 40,000 million in table 1 = 10,000 million) assumed to be business checks (see footnote 4, above), giving \$2,400 million in costs. Adding this to the \$11,031 million in all payor mailing costs, yields \$13,431 in total payor costs.

[6] Bank processing/transportation costs per check deposited were \$.11 per (weight unit) item in 1982, based on the latest *Functional Cost Analysis* report available. It was assumed that the interest and reserve requirement costs of demand deposits were offset by requiring higher average balances to be held so these costs were not included in bank check costs. Thus the 11¢ figure is demand deposit operating cost divided by a weighted average of "on-us", transit, and official checks deposited and processed for some 608 banks in 1982. This figure was reduced by 28% to reflect direct cost recovery through service charges and thus limits the potential for double counting fees paid by payors and payees and bank costs. The result, 8¢ per check, was multiplied by the estimated 40 billion checks written in 1983, giving total bank check costs (net of double counting) of \$3,200 million.

[7] All Federal Reserve costs were taken from 1983 PACS data (*Summary Reports*) which included both direct costs and an overhead cost allocation. All costs were raised by 16% (except cash, which is largely a nonpriced service) to reflect the effect of the PSAF (or private sector adjustment factor) then in place.

[8] Total check float in 1983 has been estimated to be \$380 billion on a daily average basis (Dudley, 1983, p. 11). Evaluated at the 1983 average 90 day Treasury bill rate, this gives a total check float cost of \$33 billion a year. This cost is composed of mail float (\$11 billion), payee delay in depositing checks (\$17 billion), delays in bank check processing and collection (\$5 billion), and Federal Reserve float (\$120 million). Prior to the passage of the MCA in 1980, Federal Reserve float costs rose to a high of \$655 million in 1979 (or \$6.5 billion in float evaluated at 10.07%). As seen, the Federal Reserve float component is small relative to the other components.

[9] Payor ACH costs are estimated to be \$.18 per ACH item, based on the government's direct deposit ACH costs (Dudley, 1984). Applying this origination cost to the total volume of ACH items originated (400 million items, table 1), gives \$72 million as an estimate of government and business payor costs. Payee and bank ACH costs were derived by multiplying the 1981 average per item commercial bank price for ACH (see table 3, below) by 1983 ACH item volume (see table 1). The bank cost of handling ACH tapes was derived by multiplying the average per tape fee in table 3 by the estimated number of tapes handled at Reserve Banks. Reserve Banks are estimated to have handled some 116,000 commercial and 9,000 government ACH tapes in 1983, for a total of 125,000 tapes per year. There are many more ACH items on a government tape than typically occur on a commercial tape, a fact easily seen when it is noted that while 61% of the ACH volume in 1983 was government originated, they accounted for only 36% of the tapes deposited. The fact that 1981 commercial bank prices were used means that Federal Reserve costs are only partly reflected in these fees (because Federal Reserve prices recovered only 20% of Reserve Bank ACH costs in 1981, the first year of ACH pricing). Actual Federal Reserve costs were \$44 million in 1983 (including the PSAF) while private sector costs were estimated to be \$21 million (\$.046 per item × 400 million ACH items + 125,000 tapes × \$19.15 per tape = \$21 million).

[10] ACH debits do create some float (\$68 million in 1983), like checks, but there is no float associated with ACH credits (which are like European giro payments). ACH debit float was evaluated at the 90 day Treasury bill rate cited above, giving \$6 million in

nents. First, the transfer payment component of the two most widely used payment methods—cash and checks—exceeds the real resource costs involved. The important difference, however, is that cash users *pay* this transfer payment (to the government) while check users *receive* the transfer payment (from payees). In the first case user costs are raised while in the second they are reduced. Second, the total real resource or social costs of cash and checks far exceed that for ACH and wire transfers. This is expected because cash and checks account for 99% of all payments transactions (table 1). Lastly, the unit social cost of the different payment methods (total social cost divided by total volume in table 1) seemingly favors cash or ACH electronic payments versus paper checks but the unit user costs continue to favor checks. In fact, the float benefits of checks are so large that they exceed the real resource costs involved so that the direct user check cost is *negative* (see next to last row in table 2). These and other results are now discussed in more detail.[10]

The real resource costs shown in table 2 include all production and processing costs but only account for 42% of the total expenditures made to use cash and checks. There are considerable windfall gains to check users from check float and to the government from the opportunity cost of idle funds. Float gain to check users is based on the fact that a check is effectively a sight draft subject to signature verification prior to payment by the payor institution. Over the 100 years that checks have been in existence, it is very likely (but unverifiable) that the gain in float from check usage by payors has been partially (or completely) offset by price increases by payees, at least in the aggregate. This is likely because similar responses by payees have been observed in response to changes in sales or corporate income taxes. That is, although a float "tax" may be levied on payees who

---

[10] The costs shown in table 2 are estimates or approximations and not everyone will agree with how they have been derived. The derivations are shown in the notes to the table. Such cost calculations have only rarely been attempted in the past and they all, including this one, suffer from a lack of good data. Gambs (1981) represents one of the better earlier efforts in this area.

---

float. Wire transfer also can create some small amount of float as well, only in this case it is typically credit float, not debit float ($55 million for FedWire in 1983, none for CHIPS). Credit float is created when Reserve Banks debit a sender's reserve account but, because of operational difficulties, do not pass credit to receivers until the next day. Wire transfer credit float is thus a user cost, not a benefit like check or ACH debit float.

[11] Wire transfer volume in table 1 was multiplied by the average commercial bank charge (table 3) for wire transfers to give the total cost of wire transfers ($297 million). The Federal Reserve wire transfer cost component, from PACS plus the PSAF, was $58 million and was assumed to be fully passed on to final users in bank fees. Thus the payor, payee, and bank cost component of wire transfers was determined as a residual ($239 million).

[12] User and social unit costs were all derived by dividing the total user and social costs in this table by the volume figures in table 1. The result is a per transaction cost estimate.

elect to accept checks rather than restrict all payments to cash, ACH, or wire transfers (the "no-float" payment alternatives), payees can pass that tax forward to payors. This is done by raising prices to all payors regardless of the payment method used or by some extra charge (discount) for payors who use checks (cash or electronic payments). The latter method is administratively more complex but is occasionally used today. Some recent examples would be cash discounts at certain gasoline stations if a credit card is not used and nonprice barriers (minimum purchase restrictions) for acceptance of a check or credit card for payment. How much of the payee float "tax" has been passed forward to payors ultimately depends, of course, on the elasticities of demand and supply for the goods being purchased. This issue—the incidence of a tax—is an old (and complex) problem in public finance and basically boils down to the relative market power of payors and payees. For example, when especially large dollar transactions are being discussed, it is common to negotiate the method and timing of payment. If the payor is in a strong bargaining position, a check may be used; if the payee has the advantage, a wire transfer is likely to be required. For small dollar transactions, such negotiation would be expensive and the cost of check float to payees is effectively folded into the price of the service or product being provided.

The main point requiring emphasis here is not that a payor's check float has almost certainly been largely offset by payee price increases over time. The important point is *because* these price increases have usually not discriminated between types of payment methods, that noncheck payments are inherently more expensive since no float is obtained by payors to offset the price rise instituted by payees. Thus from the user's perspective incentives will be necessary to induce check users to switch to an alternative payment method (such as ACH or debit card). The influence of check float on relative user costs and the problem it creates for efficient use of real resources can be illustrated by an analysis of the government's direct deposit program. This is where government paychecks, social security payments, and retirement checks are directly deposited electronically through an ACH rather than printed, processed, and mailed to recipients.

In the most recent and sophisticated analysis of the direct deposit program, Dudley (1984) has found that the real resource cost of an ACH direct deposit payment is 27¢ while it is 40¢ per payment if a check based system were used. But for each transfer made by ACH direct deposit, the government gives up 66¢ in check float. In terms of user cost, which includes the benefit of check float, the government price for using check disbursement is −26¢ (from 40¢ in real resource costs less 66¢ in check float benefits). This negative cost per check contrasts unfavorably with the positive cost of 27¢ per ACH direct deposit. The cost to the user clearly favors a continued use of checks even though real resources could be saved by switching to ACH payments, since the real costs of ACH are 27¢ while each check costs 40¢ in real resources. This mirrors the unit cost result reported at the bottom of table 2 for all checks. As a user of payment services, the government can save tax revenues by using checks but the nation as a whole loses on each item *not* converted to the ACH direct deposit program.

Should the government act as a private sector user would and minimize user cost, which is here dominated by the effects of a float transfer payment, or should it move to maximize output of the nation as a whole by minimizing social cost? The choice is made more difficult by the knowledge that in minimizing social costs, the government usually will have to raise taxes to account for the fact that the real resource savings obtained does not typically accrue solely to the government sector but rather to some other segment of society.

While such user-society conflicts may appear to be rare, they are disconcertingly abundant among the payment mechanism issues raised below. Similar private-social concerns are involved in:

- Determining how best to price and/or eliminate Federal Reserve float, as mandated by the Monetary Control Act;
- Assessing the value of providing immediate and non-reversible funds availability (payments finality) of large-dollar wire transfers over FedWire;
- Developing a daylight overdraft/exposure policy for all large-dollar wire transfer networks;
- Allowing (conditional) access to reserve accounts for net settlement; and
- Pricing of and access to discount window loans.

All but the last issue is discussed in detail below.

B. Commercial Bank Prices for Payments Services

Commercial bank prices for payments services are shown in table 3. These fees apply to correspondent bank and corporate users, not to the typical consumer. The typical consumer has never heard of (much less used) a wire transfer; he is very unlikely to know what an ACH is even if his paycheck or social security payment is directly deposited; and he pays for check (and automated teller machine or cash dispenser) services typically through holding small idle or larger interest paying balances rather than through direct fees (other than check printing costs). Bank earnings on these balances, even after netting out the operating costs of supplying "free" deposit services, are said to provide income spreads of from one to two hundred basis points greater than spreads available with purchased funds (CDs, federal funds, Eurodollars). These deposits are also a more stable source of funds than are purchased money. This has been demonstrated in the banking literature, and more recently, at Continental Illinois.

TABLE 3

Commercial Bank Fees for Payment Services Provided
to Respondent Banks
(1981)

	Average	Low	High
Cash (per $1,000 furnished)	$.42[1]	$.03	$.75
Checks (per encoded item)	.04[2]	.012	.12
ACH (per item plus per tape)	.046 (item)	.01	.10 (item)
	19.15 (tape)[3]	10.00	20.00 (tape)
Wire (regular transfer)	5.21[4]	.80	10.00

Source: Robert Knight Associates, 1981. There were 151 banks in this survey of commercial bank correspondent fees.

[1] Neglects other fees, such as a $2 to $20 fee for each currency cash order. Coin orders have a higher price.

[2] Most checks are encoded prior to being deposited at a collecting bank. The prices shown are for checks drawn on banks in the same Federal Reserve zone and cleared through correspondents.

[3] Listed under "Account Analysis for Selected Corporate Banking Services" from the same source. ACH credits are priced somewhat differently from ACH debits.

[4] Different types of wire transfers—FedWire, BankWire, CHIPS—are similarly priced by banks.

The prices shown in table 3 are averages of prices for 151 commercial banks in a yearly survey conducted by Robert Knight Associates. The low and high prices observed are also shown in table 3 and clearly indicate that there is substantial variance across banks in the prices that are charged. This variance, which has persisted since the first survey was published in the early 1970s, reflects cost differences and/or differing degrees of competitiveness across banks in offering this service in different market areas.

The payment service fees shown reflect the processing costs incurred by banks and those charged by the Federal Reserve to banks. These costs, in turn, are passed forward to payees and payors and, as a whole, comprise one part of the real resource processing costs shown in table 2. While prices are shown in table 3, the charging method used is often in terms of compensating balances, not direct fees.

C. Direct Fees Versus Balance Requirements

In charging for payments services, most banks use balance requirements—an indirect charge—rather than explicitly levying a direct fee. In fact, over 90% of the bank cost of checks and wire transfers are paid for through balances rather than fees (Robert Knight Associates, 1981). This holds for payments services used by consumers, corporations, as well as respondent banks. And, until the passage of the Monetary Control Act, reserve requirement balances were in effect used to pay for payment services provided by Reserve Banks.

In principle, a direct fee should be a perfect substitute for a balance requirement since both can be so structured to capture an identical amount of revenue for the payment services provided. But if this is so then one would not *a priori* expect to see such a current and historical dominance of balance requirements versus direct fees *across* different geographically isolated banking markets. The historical reasons for preferring balance requirements as a charging method seem to be:

(1) To Facilitate a Funds Transfer: balances serve a dual purpose since the interest they earn is used to pay for payments services provided and they simultaneously serve as a source of credits against which debit transactions can be functioned;

(2) To Promote Interbank Liquidity: balances are a liquid source of funds for depositing respondent banks which can be drawn down to cover large and unexpected deposit withdrawals; and

(3) To Provide a Stable and Low Cost Source of Funds: balances used to pay for payments services are a more stable and reputedly lower cost method of funding assets for correspondent banks than are purchased funds, particularly when excess balances are held (as often occurs in practice).

The first reason relates to the fact that reciprocal balances were needed to settle interbank funds transfers prior to the establishment of the Federal Reserve. Use of reciprocal balances was less expensive and more secure than settlement through the constant movement of gold between bank vaults. Check users paid for payment services through discounts to the face or par value of the check being deposited. Actually, this was very close to a direct fee since, at that time, the prior method of settlement (moving gold between bank vaults) *was* directly related to the value of a check. This relationship, although more tenuous, continued when reciprocal balances were used for settlement since more balances are drawn down when large funds transfers were made.

Nonpar banking, as practiced, collapsed what would have been two transactions into one. Checks could have been paid at par and then the depositor charged a direct fee based on the par value of the item deposited. It was simplier to discount the par value of the check and retain a larger reciprocal balance. Since these practices existed prior to the development of any significant market in purchased funds, it was natural for banks to emphasize balances as a charging/clearing method over direct fees. Lacking a sophisticated purchased funds market, direct fees were not as commonly used to fund assets as they are today, where fee income can finance liability purchases for funding new loans.

While the first reason for preferring balances over direct fees appears to have benefitted both the balance owner and the balance holder (so that idle balances required for settlement purposes could be folded into balances used to pay for services), the second (third) reason primarily benefits the balance owner (holder). While these three reasons would seem to be sufficient to explain the preferences for balances over direct fees, there are other (weaker) reasons which have been offered:

(4) Prohibition of interest payments for demand deposits and interbank deposits (so-called banker's balances);
(5) Tax effects: no tax on bank services provided as "income in kind"; and
(6) Marketing: easier to "hide" the true cost of payment services when balance requirements are used.

The fourth reason is a recognition of the role played by current legislation, passed in 1933. By prohibiting interest payments on balances held in accounts most directly related to customer and interbank transactions, legislation has prevented banks from directly compensating users for posting (idle) balances needed for settlement of the transactions and funds transfers which the user wishes to make. These idle balance costs are minimized, however, if users post balances to pay for the payments services they desire, since these balances can also be used in settlement. Consequently, balance requirements, but not direct fees, serve to reduce the cost of settlement to all parties to a funds transfer. If these interest prohibitions were eliminated, then interest could be paid on clearing/settlement balances and direct fees charged for payments services without raising the cost of making a funds transfer.

The tax effect noted in (5) refers to the fact that state and federal income taxes do not apply to "free" bank services provided to consumers in exchange for balance requirements whereas direct interest payments add to a consumer's taxable income base—and are not offset by the expense of direct fees for payment services (except for businesses). The last reason (6) concerns the apparent ability to mask or "hide" the true cost of payment services to certain classes of users when a balance requirement is used compared to direct fee. For sophisticated users, a balance requirement is easily translated into its equivalent direct fee and comparison shopping for the lowest equivalent direct fee is typically undertaken. These users generally are corporations. For the typical consumer user, however, accessibility of a branch office is usually the main concern in choosing a payments service supplier and very little "rate shopping" occurs.

The preference for balances in order to "hide" the true cost of banking services is not unusual. Check float costs are themselves "hidden" in the (higher) price of retail services for the typical consumer. What the check user sees as a float benefit is in fact paid for by him and users of cash in retail transactions. Since check users do not see that they are being (indirectly) charged for float and, more importantly, do not see any retail price reduction if they use cash or an ACH transfer to pay for these services, they continue to see float as a benefit of using checks.

IV. The Pricing of Federal Reserve Payments Services

A. The Quid Pro Quo of Federal Reserve Pricing

The Federal Reserve Membership "Problem," Deposit Coverage Ratio, and Monetary Control. The primary impetus for the Depository Institutions Deregulation and Monetary Control Act of 1980 (MCA) was not to price Federal Reserve payments services but, as the title of the Act implies, to deregulate interest rates and preserve monetary control. With respect to monetary control, it was argued at the time that transactions deposits subject to reserve requirements were falling as a percent of total transaction deposits and this, if it continued, would make some aspects of monetary control more difficult. It would not have made monetary control under an interest rate target more difficult (the policy followed when the MCA was first proposed) but it limited the ability to effectively use a money supply target. Monetary control under a total reserves or money supply target would be more difficult because the existence of banks without reserves increases the unexpected variation in the money multiplier which relates reserves and money supply to aggregate economic activity.

The transaction deposit coverage ratio was falling because Federal Reserve (FR) member banks were withdrawing from membership in order to escape reserve requirements and because deposits at nonmember banks grew at a more rapid rate than at member banks.[11] Federal Reserve member banks obtained free payment services, which they used themselves and resold to nonmember depository institutions and corporations as correspondent bank services. Member banks also held idle required reserve balances at Reserve Banks. The lost earnings on idle reserve balances, however, was estimated to exceed, by some $700 million in 1978, the benefit of (a) free FR payment services, (b) FR float, and (c) the imputed benefit of discount window access.[12] This burden on member banks, of course, affected their profitability and provided a strong incentive to withdraw from membership. It was estimated that non-member banks, had they been subject to Federal Reserve reserve requirements, would have seen their net incomes fall by 9 to 17 percent, with the largest banks experiencing the greatest decline.[13]

---

[11] In the late 1960s and early 1970s, the more rapid growth rate for nonmember bank deposits accounted for about 80 percent of the reduction in the deposit coverage ratio (Rose, 1979).

[12] This estimate represents an internal FR staff update and revision of the FR membership burden estimate made earlier in Paulus, 1977. This issue has been comprehensively reviewed in Rose and Rose, 1979.

In order to eliminate the incentive to withdraw from FR membership, an action which also decreased the percent of total bank deposits subject to reserve requirements and complicated monetary control under a money supply targeting procedure, it was suggested that interest be paid on reserves. But with some $28 billion held as idle balances at Reserve Banks in 1978, Congress favored a less expensive alternative and chose instead to:

(1) Lower reserve requirements on all FR member banks over 4 years;
(2) Impose reserve requirements on all nonmember banks and depository institutions holding transactions balances over 8 years (S&Ls, credit unions, MSBs);
(3) Price all Federal Reserve payment services; and
(4) Price and/or eliminate all FR float.

Lowering reserve requirements on member banks would reduce the "tax" on these institutions but pricing of previously free FR payments services and pricing or eliminating FR float was viewed as the *quid pro quo* of this benefit.

Finally, the imposition of equal reserve requirements on all depository institutions—which will be fully phased-in by 1988—both equalized the reserve "tax burden" between member and nonmember institutions and stabilized the ratio of reservable transaction deposits to total transaction deposits in the U.S. banking system, which previously had been falling. This ratio, which was 86 percent in 1950, dropped to 72 percent in 1978—it is now 67 percent,[14] and was forecasted by FR staff to fall to 57 percent in 1985 if the MCA had not been passed.[15] Although a number of academics writing on this subject[16] or who testified before Congress in hearings on the MCA suggested that reserve requirements were not absolutely necessary to conduct an interest rate target monetary policy—that is, such a policy could be conducted solely by using open market operations and altering the discount rate—the Federal Reserve maintained that having reserve requirements as an additional tool made the conduct of monetary policy easier and more effective. In addition, imposing equal reserve requirements across all depository institutions would serve to reduce the unpredictable component of the money multiplier and so better stabilize the relationship between reserves and aggregate economic activity under a total reserves or money supply operating target.[17] Although the Federal Reserve was basically following an interest rate target six months prior to the passage of the MCA in early 1980, it is doubtful that it could have as easily maintained the switch to the money supply targeting procedure in October 1979 without having all banks subject to reserve requirements at some point after the switch.

---

[13] This estimate takes into account that most states have reserve requirements as well, although often these state reserve balances can be held in earning assets (Goldberg and Rose, 1976).

[14] The ratio has fallen because reserve requirements at member banks are phased down twice as fast as they are being phased up for nonmembers. The continued phase-up for nonmembers through 1988 after the member bank phase-down is completed in 1984 will raise the coverage ratio.

[15] This ratio is not 100 percent because 62 percent of all depository institutions—basically the smaller institutions—hold idle vault cash which exceeds what their reserve requirements would be and so, in effect, are not "bound" by the requirements. That is, their deposits can fluctuate without affecting reserves held at Reserve Banks.

[16] See Benston, 1978 or Starleaf, 1975.

While Congress was persuaded to grant the Federal Reserve additional flexibility in pursuing monetary control, making the switch to money supply targeting from interest rate targeting more feasible, it was also concerned about revenues. Reserve requirements were viewed as an important "user fee" or "franchise tax" on the banking system which, if eliminated, would have to be made up out of general revenues.

Increased Treasury Revenues from the Monetary Control Act. While payment of interest on reserves can clearly reduce Treasury revenues (by $2.2 billion a year if a 10 percent interest rate were paid on the approximately $22 billion in reserves at Reserve Banks today), the MCA actually is increasing Treasury revenues compared to what is estimated would have occurred without the MCA. Indeed, the Federal Reserve agreed to compensate the Treasury for up to $1 billion from their capital and surplus accounts if any cumulative negative effect on Treasury revenues were experienced from passage of the MCA. Over 1981-1988, which corresponds to the time period that non-member depository institutions will be given to phase-in reserve requirements fully, the total net (after tax) gain in Treasury revenues from passage of the MCA is estimated to be $4.4 billion, or some $550 million a year. The main elements in this computation are shown in table 4.[18] The primary reason why the Treasury gains, rather than loses, from the MCA is that this Act stopped the attrition in FR membership—and lost revenues from reductions in idle reserve balances—which otherwise would have occurred. This reason, by itself, accounts for some 75 percent of the net gain in Treasury revenues had the MCA not been passed.

B. Federal Reserve Pricing and the DIDC: Two Forces for Change In Payments Practices

The Cost-Push and Demand-Pull Effects of Federal Reserve Pricing. The pricing of Federal Reserve services was required

---

[17] See Kopecky, Parke, and Porter, 1983 for a detailed analysis of this issue.

[18] This internal FR staff analysis is much too complicated to be discussed here but is similar in concept to the earlier analysis published in the *Congressional Record - Senate*, March 27, 1980 (S-3172-74).

TABLE 4

The Monetary Control Act and Treasury Revenues[1]
($ millions, calendar years)

	1981	1982	1983	1984	1985	1986	1987	1988
1. Earnings Loss from Reserves Released (reduced reserves on FR members, increased reserves on nonmembers)	−51	−186	−480	−763	−769	−605	−455	−284
2. Revenues from Charging for FR Payment Services	154	422	529	650	717	792	869	955
3. Earnings Gain from Float Reduction	195	309	449	507	546	587	617	647
4. Revenues from Charging for Float	0	0	172	162	162	162	162	162
5. Cost Reduction Due to Volume Reduction	28	37	72	82	92	104	116	129
6. Revenue Shortfall (without PSAF)	−12	−35	0	0	0	0	0	0
7. Total Change in Gross Treasury Revenues	314	548	842	739	849	1040	1309	1609
8. Offset for Tax Revenues on Change in Bank Earnings	−100	−186	−254	−231	−280	−388	−488	−601
9. Total Change in Net Treasury Revenue	214	362	487	407	468	652	821	1008
10. Bank Earnings on Reduced Reserve Requirements (line 1) as a Percent of FR Payments Service and Float Costs (lines 2, 3, and 4)	15%	25%	42%	58%	54%	39%	28%	16%

Source: Internal Federal Reserve staff estimates.

[1] 1981 and 1982 are actual data; data for 1983 and beyond are projections.

by the MCA and has increased the cost of all payments services: using coin and currency, writing checks, completing an ACH transaction, and making a wire transfer. These costs were increased directly—cost-push—for users of FR services and have given alternative suppliers of payment services the opportunity to raise their prices as well. Such alternative suppliers would be correspondent banks and data service bureaus which provide coin, currency,[19] and check processing services and those large banks which collectively own large-dollar wire transfer systems. Alternative suppliers of payment services have the opportunity to raise prices and not lose volume since, in theory at least, the establishment of non-zero Federal Reserve prices will have led to an outward shift in the demand curve—demand-pull—for alternatively supplied payment services at all prices. The price of alternatively supplied payments services could remain stable, so that the shift in demand leads to a volume increase, or it could rise and reduce the volume increase experienced.

In check processing, yet another influence was felt—the entry of new suppliers. The pricing of Federal Reserve services led to the new entry or expansion of 95 local clearinghouse associations which took check processing volume away from *both* of the two predominant suppliers—the Federal Reserve and large correspondent banks. These new local clearinghouse entrants were concentrated in the eastern half of the U.S., as can be seen in Figure 1, where distances between economic centers is shorter and average population density is higher. Perhaps not unsurprisingly, this was the same geographic area which experienced a reduction in local check clearinghouse operations when the Federal Reserve introduced in the early 1970s improved availability for certain checks at existing offices and new branch offices as well—otherwise known as Regional Check Processing Centers (RCPCs)—at a zero (direct) price to users. Many of the existing local clearinghouses went out of business at that time only to re-emerge with the implementation of FR pricing.

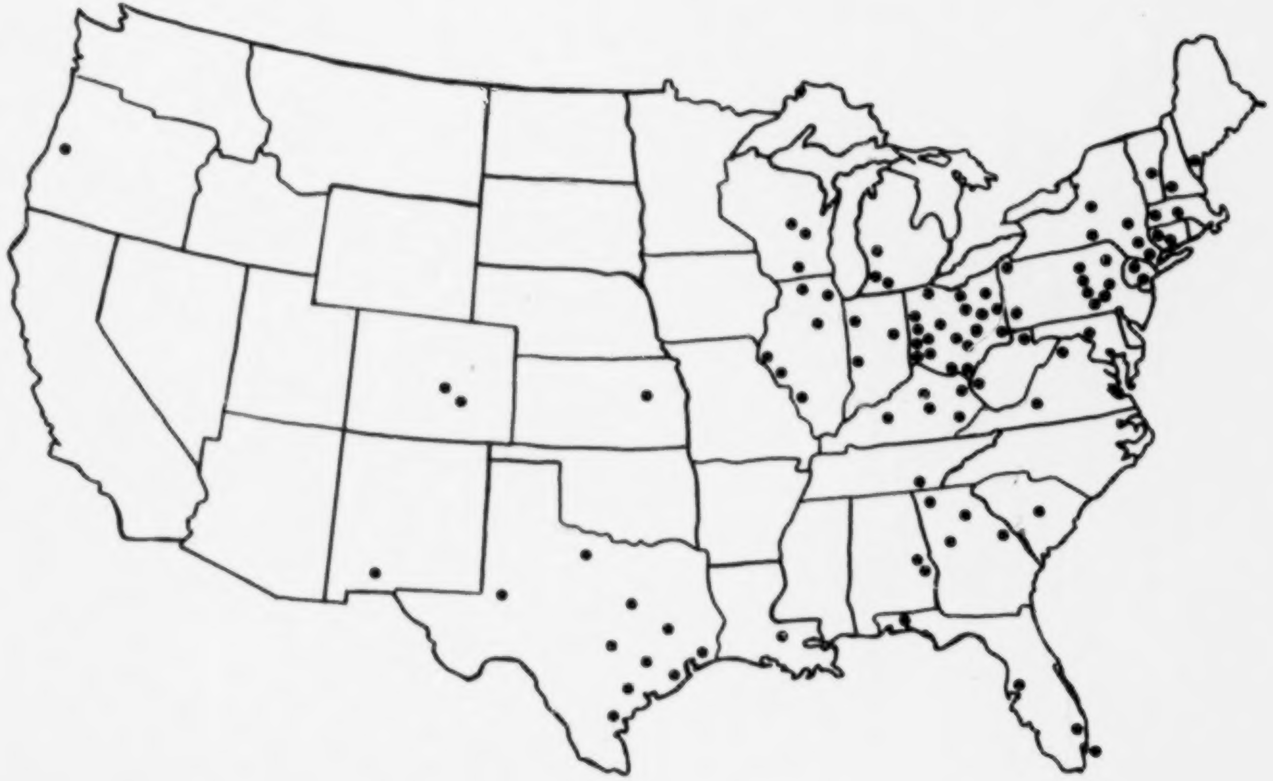
Overall, FR pricing will raise the payments processing costs of depository institutions who, in turn, repackage these wholesale operations into payments products for their retail consumer and corporate customers. But, as noted in the previous section, the impact of FR pricing has been partly offset, *on average*, by reductions in reserve requirements. This can be seen in line 10 of table 4 where the bank revenues obtained from reduced reserve requirements (line 1) are divided by the total costs of FR services (line 2) and FR float (lines 3 and 4). In 1984, this offset was 58 percent, which represents the

---

[19] Most of the coin and currency services are still provided free as a central bank service. These services include minting of coin, printing of currency, shipment to Reserve Banks, processing for counterfeit bills, and destruction of unfit currency.

FIGURE 1

New or Renewed Local Check Clearinghouses Activated Since FR Pricing of Payment Services in 1981
(95 dots show areas of new or renewed activity as of 1983)



Source: Adapted from Frodin, 1984.

maximum, so that on balance the cost-push effect of FR service pricing and float reduction is less than would otherwise be the case. Over the entire eight year period, 36 percent of the cost increase to all banks from FR pricing and float reduction has been offset by reductions in reserve requirements.

If we divide bank customers into consumer and corporate groups, it is clear that only the latter has the sophistication and market power to "persuade" banks to limit their price increase to the net effect of the FR service price, float reduction, and reserve requirement reduction effect rather than pay the full cost-push effect of FR pricing and float reduction alone. And we would naturally expect banks to pass on their cost increases according to their ability to segment their market into those customers which have inelastic demands (individual consumers and small businesses) versus those customers with elastic demand (corporate customers). Such actions are expected and termed Ramsey pricing in the economics literature. And since such pricing behavior is to be expected from the private sector, one should not be surprised when it is adopted under the name of market sensitive pricing when applied to the pricing of FR payment services, as discussed below.

Federal Reserve Prices for Payment Services. Prices for selected payments services are shown in table 5. These prices are national averages, masking the differences which actually occur in different regions in the U.S. Coin and currency delivery services are actually supplied by armored carrier firms, such as Brinks, Wells Fargo, Purolator, and other more local firms, which bid to supply these services through the Federal Reserve to end users. The competition here is primarily between the Federal Reserve contracting armored carrier and other carriers (or even the same carrier) who often are able to undercut the FR contract price by relying on marginal cost pricing and the existence of conjunctive business to coin and currency delivery points close to those serviced by Reserve Banks. Since the Federal Reserve effectively serves as a middleman for this payments service, Reserve Banks have at times been stuck with coin and currency delivery contracts that can not be altered rapidly enough to offset the decline in delivery endpoints served. A solution to this problem has merely been to arrange for coin and currency delivery between groups of users and armored carriers rather than be a party to any contract for delivery, which was common practice prior to the MCA and the pricing of this transportation service.

In check processing, price is only one of the important determinants of choosing which supplier to use. The other important determinant is funds availability. Typically, correspondent banks offer

TABLE 5

Federal Reserve Prices for Payment Services
(National averages, 1984)

	Average		Low	High
Cash Services:				
Currency Delivery (per $1,000 shipped)	$.17		$.09	$.57
Check Services:				
Fine Sorted Checks	.014		.010	.018
Regular Processed Checks	.029		.022	.031
ACH Services:				
Processing ACH Items (orgination fees)	.03	(debit item)	National prices	
	.01	(credit item)		
	3.00	(per tape)		
Wire Transfer Services:				
On Line Transfers	.65		National prices	
Off Line Transfers (orgination fees)	6.15			

Sources: Federal Reserve prices are (unweighted) averages across Federal Reserve Districts for cash and check services and, at times, across different types of specific services provided. ACH and wire fees are national prices and therefore the low and the high values are the same as the average. Data are from internal Federal Reserve sources and press releases.

better funds availability but can only do so because more specialized and higher cost transportation and processing services are used to provide it. Thus, when the value of speedier funds availability offsets the higher processing and transportation costs of correspondent banks, which often occurs for non-local checks larger than $300, users will deposit these items with a correspondent network rather than the Federal Reserve. For low value checks, the opposite is true.

In the wire transfer area, the services provided by the Federal Reserve and groups of correspondent banks are very similar. The main difference lies in the fact that only in one of the two primary U.S. large dollar funds transfer networks—FedWire and CHIPS—is it possible to treat the funds received as final or "good" funds during the day. Funds received over CHIPS—a network owned by large banks and operated by the New York Clearing House Association—are treated as provisional funds until settlement occurs at the end of the day, which then makes the transfer final. In contrast, funds received over the Federal Reserve's FedWire network are final when made since Reserve Banks provide a 100 percent guarantee for all funds received even if the sending institution should fail without adequate funds in its reserve account. However, the assumption of this risk of loss by Reserve Banks has not been incorporated into the price of the wire transfer service and, in this sense, the price is understated.

The DIDC and the Pricing of Payments Services. The Depository Institutions Deregulation Committee (DIDC) was set up by the MCA to phase-out by 1986 regulatory ceilings placed on interest rates depository institutions could pay their depositors. This phase-out does not apply to the statutory prohibition against payment of interest on demand deposits; it does apply however to interest ceilings on all other types of deposits (small time deposits, savings deposits, large time deposits—CDs, share draft deposits, and deposits in NOW accounts). This charge to the DIDC has now been almost fully realized and affects the pricing of payments services as follows:

(1) Deposits pay interest at close to market rates (except for demand deposits);
(2) This has reduced the earnings spread on these liabilities;
(3) To cover the operating costs of servicing these deposit accounts and expand the earnings spread, depository institutions have responded by:
    (a) Setting higher maximum balance requirements (so that the smaller earnings spread is offset with a larger average balance and total earnings rise while operating cost per deposit dollar falls); and
    (b) Imposing direct fees for servicing these accounts to cover the operating costs previously absorbed when earnings spreads were greater.

At this writing, most depository institutions are both imposing some direct fees and setting higher mimimum balance requirements in response to the deposit rate deregulation. To the extent that direct fees are assessed for the type and/or number of account transactions made (e.g., a fee per account withdrawal or check written) or balance requirements are tied to the transactions made, the DIDC actions will raise the perceived cost of payments services to consumers. In addition, depository institutions face increased costs in providing check, ACH, and wire transfer services as a result of FR pricing. In response, consumers will be induced to:

(1) Consolidate savings and small time deposit accounts with interest paying transaction accounts (such as NOWs, Super NOWs, and share draft accounts) to reduce the overall service cost and balance requirements of deposit services; and

(2) Choose alternative payment arrangements which may be less expensive and/or more convenient to use, such as using ATMs or cash dispensers or using debit cards to obtain funds to pay for purchases rather than write checks.

Thus, on balance, consumers should over time reduce the quantity of traditional deposit/payment services demanded and increase use of substitute payments methods, such as cash and perhaps credit/debit cards (when available). Depository institutions, because of greater pressure on earnings spreads and actual operating cost increases from FR pricing, will seek to minimize the rise in operating costs by providing lower cost alternatives, the most important of which are customer use of ATMs for servicing deposit accounts and providing cash in lieu of continued reliance on traditional teller and brick and mortar payment service delivery methods. Overall, FR pricing and DIDC actions have served to pass the true costs of payments services on to the ultimate users (consumers) and because of this, permanent changes in both the types of payments services offered and how they are delivered to users are occurring and will continue in the future.

C. Choice of a Pricing Strategy for Federal Reserve Payments Services

How should the prices of Federal Reserve check, ACH, wire transfer, and other services be determined? In practice the Federal Reserve has little room to maneuver regarding a pricing strategy because the MCA has mandated that prices cover "all direct and indirect costs actually incurred," including float, overhead, and costs that would have been paid by a private provider (debt costs, all taxes, return on equity capital, and the like).[20] This pricing guideline was followed in the schedules implemented in January 1981 for wire transfer services and in August 1981 for check and ACH services. Prices were *initially* set equal to their imputed average costs based on Reserve Bank allocations of fixed and variable costs to individual priced services. These prices incorporated a private sector adjustment factor (PSAF) or "mark up" of 16 percent.

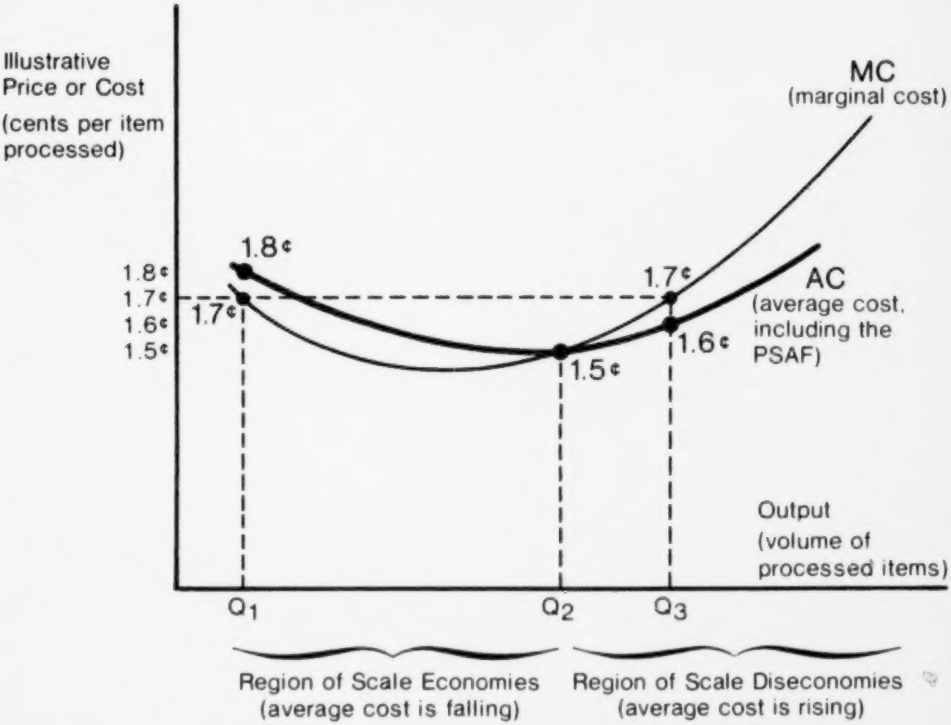
---

[20] See *Depository Institutions Deregulation and Monetary Control Act of 1980*, S. Rept. 96-640, 96 Cong. 1 Sess, Government Printing Office, Washington, D.C. (1980).

This pricing procedure is illustrated in figure 2. If the check output level being produced is at Q_1, Q_2, or Q_3, the average cost (which here includes the PSAF) assumed to be associated with these output levels is 1.8¢, 1.5¢, and 1.6¢ on the average cost curve shown in the figure. Depending on the level of output produced, these would be the prices charged with average cost pricing. Average cost pricing was a convenient way to develop prices for services which were never before priced and provided a simple way to implement detailed cost allocation procedures which had never before been required. It also provided a way to attempt to achieve a cost/revenue match without developing complicated pricing guidelines for Reserve Bank personnel

FIGURE 2

The Relationship Between Marginal Cost, Average Cost, and Price Under Different Scale Economy Situations



who, at that time, were unfamiliar with pricing exercises. Given the legislative restrictions on pricing, which were written to ensure competition from the private sector for all priced services, the optimality of the initial pricing exercise is now examined using results on Reserve Bank payments processing scale economies (discussed below) and relevant information regarding the pricing practices of correspondent banks.

Marginal Cost Pricing. Economic theory recommends marginal cost pricing as the preferred pricing method to ensure an optimal allocation of resources. When price equals marginal cost, shown as 1.7¢, 1.5¢, and 1.7¢ for the three output levels illustrated in figure 2, users face the true real resource cost of producing the product or service in question and are thus able to make socially optimal decisions regarding the use of that service compared to others. The end result is that the mix of outputs demanded is produced at the lowest real resource cost so everyone benefits. Other pricing considerations, such as matching revenues to costs (important in public utility pricing) or operational feasibility (important in cases in which it may be difficult to measure marginal costs accurately) have been treated as secondary to resource allocation in the economics literature.[21]

Marginal cost pricing by all market participants is a "first best" solution. However, if most participants price in other ways, adoption of marginal cost pricing by only some competitors—a "second best" solution—may not improve resource allocation (Lipsey and Lancaster, 1956). Correspondent banks that provide payments services competitive with those offered by the Federal Reserve do not practice marginal cost pricing. These institutions generally view their correspondent services as a package. Losses or low profits on some services (say, check processing) are offset by higher profits on others (say, specialized management services). Pricing is structured to make the entire correspondent relationship profitable (and competitive) rather than to make profitable each product within a service line. Because private suppliers do not practice marginal cost pricing, such pricing by the Federal Reserve alone creates a second-best solution. Adoption of a different pricing strategy therefore need not worsen resource allocation. It is not obvious, however, that average cost pricing by the Federal Reserve is an optimal divergence from marginal cost pricing.

Although treated as secondary to the benefits of proper resource allocation in the economics literature, it is worth noting that the MCA directed the Federal Reserve to price in such a manner that total revenues match total costs (plus the PSAF). Marginal cost pricing, however, does not ensure that total revenues raised will match total costs incurred in production. Only in the special case where there are no technological scale economies or diseconomies (i.e., a flat average cost curve exists, shown here only at one point, Q_2 in figure 2) would marginal cost pricing ensure such a cost/revenue match. But in this instance, average cost is identical to marginal cost and so average cost pricing would also be optimal. In the more general case where marginal and average cost are not identical, (i.e., at output levels other than Q_2 in the figure) marginal cost pricing for each Federal Reserve payments service line[22] would cause service lines that experience scale economies (at output level Q_1) to run at a loss and those that experience scale diseconomies (at Q_3) to overrecover costs. As noted below, available evidence indicates that FR provided coin and currency services, check processing, ACH, and book-entry security transfer services faced U-shaped average cost curves or downward sloping curves prior to pricing so marginal cost would *not* be equal to average costs. Thus marginal cost pricing of these four payments service lines would not in the beginning result in a cost/revenue match. Only for wire transfers did it appear that a flat average cost curve existed so that marginal cost pricing would result in a cost/revenue match for this service line alone.

---

[21] Some authors have proposed that allocational efficiency be the only pricing criteria (Miller, 1977).

In sum, while there are good reasons for economists to prefer marginal cost pricing, such as the possibility of producing a mix of payments services with the lowest real resource cost, there are at least two reasons why marginal cost pricing was not in fact chosen. These reasons are:

(1) Marginal cost pricing by the Federal Reserve alone, because competing correspondent banks do not themselves use marginal cost pricing, would be a second-best solution and, as a result, the resource allocation gains associated with marginal cost pricing need not occur; and

(2) Marginal cost pricing would not lead to a cost/revenue match, as directed by the MCA, in those services where average costs were not constant (i.e., where scale economies or diseconomies existed). Evidence indicates that constant average cost curves were the exception, rather than the rule, for FR payment services prior to pricing.

Although marginal cost pricing was rejected, the Federal Reserve ultimately adopted a pricing strategy which, when scale economies exist, is considered to be an optimal divergence from marginal cost pricing.

Optimal Departures from Marginal Cost Pricing. Although hard evidence is lacking, abundant anecdotal evidence exists which suggests that correspondent banks appear to price their payment services using a variant of the inverse-elasticity rule or Ramsey pricing. It has been shown that such pricing behavior is an optimal departure from marginal cost pricing when a revenue constraint exists (Baumol and Bradford, 1970). The inverse-elasticity rule or Ramsey pricing requires that the percentage deviation of price from marginal cost for different services be inversely proportional to each service's own-price elasticity of demand. Let P be price, MC be marginal cost, and η be the own-price elasticity of demand. When economies of scale or decreasing cost conditions exist, that is when marginal cost is less than average costs, the deviation of P from MC is related to η in the following way:[23]

---

[22] Examples of service lines are checks, wire transfers, ACH, etc. Individual services within the check service line are fine sort, group sort, RCPC items, plus all the other categories of checks processed for which individual prices are charged.

$$\frac{P\text{-}MC}{P} \cong - \frac{1}{\eta}$$

In a multiservice situation, different services will face different demand elasticities so their deviation from marginal cost will differ. And in a situation where a cost/revenue match is required, those services which face the most inelastic demands (i.e., where η is a very small negative number) will have to have prices which exceed average costs as well as marginal costs. That is, not only will prices differ from marginal costs if their demand is elastic (small difference) or inelastic (larger difference) but the prices for services facing inelastic demands will generally exceed average costs while prices for services facing elastic demands will be less than average costs. In terms of figure 2, this would mean that when scale economies exist, as at Q_1, some prices can be above 1.8¢ (when demand is relatively inelastic) while others can be between 1.8¢ and 1.7¢ (when demand is relatively elastic). The extra revenues from the former group offset the losses from the latter group so that total revenues equal total costs (plus the PSAF). In this instance, however, price should never be less than marginal cost.[24]

In this manner, resource misallocations are minimized and an overall revenue constraint can be met when scale economies exist. Resource misallocation is minimized because price will exceed marginal cost most on those outputs for which demand is the most inelastic but be close to marginal cost for those outputs for which demand is most elastic. With this basis for pricing, the overall level of quantity demanded is *larger* than if average cost pricing were used and, as a result, the cost reducing benefits from scale economies are greater. The end effect is that resource allocation will be closer to its optimum (as when price equals marginal cost) than it would have been if average cost pricing were adopted and demand elasticities neglected in setting prices.[25]

---

[23] The value of η—the own-price elasticity of demand—is negative so $-1/\eta$ is, in fact, a positive number. The formula shown here is from Berlin, Cicchetti, and Gillen, 1974, appendix B.

[24] If price were less than marginal costs, then resources would be "wasted" since the benefit to the purchaser, as measured by price, could be less than the real resource costs involved, as measured by marginal costs. Such pricing is also called predatory pricing and has often been found to be illegal by the courts. In these cases, average variable costs are typically used to approximate marginal costs, which are very difficult to measure accurately.

The same analysis applies when increasing cost conditions or scale diseconomies occur. When marginal cost is greater than average cost, MC becomes an upper bound not to be exceeded (rather than a lower bound as before). In effect, the $-1/\eta$ term in the formula above changes to $1/\eta$ so that the service facing the most inelastic demand has the price which is the most below its marginal cost (i.e., below 1.7¢, and perhaps 1.6¢ as well, in figure 2 at Q_3). Although all prices are at or below marginal cost, some will be above average costs (AC) while others will lie below it. Prices should be set in such a manner that the extra revenue from services facing elastic demands (where $P > AC$) offset the losses from services facing inelastic demands (where $P < AC$). Resource misallocations are minimized with a revenue constraint when scale diseconomies occur (as at Q_3 in figure 2) because the overall level of quantity demanded is *lower* than if average cost pricing were used. As a result, the production of services with cost raising scale diseconomies is less than it would otherwise be.

A variant of the inverse-elasticity rule, which seems particularly applicable to correspondent banks, considers not only the own-demand elasticities but the service's cross-price elasticity of demand with other services offered by the same institution and the cross-price elasticities between different suppliers of the same service (an effect influenced by the market share of the institution). Although explicit formulas exist (c.f., Sherman and George, 1979), it is very doubtful they are explicitly used to determine prices at correspondent banks. However, even casual conversations with bankers make it clear that these effects of own and cross-price elasticities of demand are implicitly considered at large, aggressive, money center correspondent banks in determining a market strategy for correspondent services. Inasmuch as these institutions are the main competitors of the Federal Reserve in supplying payments services—particularly check-processing services, for which demand is elastic—adoption of average cost pricing by the Federal Reserve, which would not consider demand elasticities in setting prices, can lead to a less than second-best allocation of resources.

---

[25] The inverse-elasticity rule has been widely used to justify the pricing practices of the U.S. Postal Service where extra revenues raised on first class mail (inelastic demand) are used to offset losses on second and third class mail (elastic demand). See, for example, Sherman and George, 1979.

Federal Reserve Market Sensitive Pricing. After one year of experience with pricing, the Federal Reserve in 1982 moved from average cost pricing for each of its services and adopted market sensitive pricing. When scale economies exist, market sensitive pricing is the same as using an inverse-elasticity rule under a revenue constraint. In this situation, Federal Reserve pricing represents an optimal divergence from marginal cost pricing because the reduction in output is being minimized. But when scale diseconomies exist, market sensitive pricing, which minimizes the reduction in output, is not optimal. As seen above, the existence of scale diseconomies reverses the procedures by which prices are to be set compared to the situation where scale economies occur and the reduction in output is to be maximized, at least until the scale diseconomies are eliminated and constant costs are attained. Thus the application of pricing procedures which are optimal for services facing scale economies are not optimal if applied unchanged to services facing diseconomies.

On the basis of the evidence regarding scale economies prior to pricing, it would appear that, by applying the same market sensitive pricing approach to all its payments services, the Federal Reserve adopted price signals which can misallocate resources in the check processing and coin and currency services (where scale diseconomies seemed to exist) but adopted prices which can optimally allocate resources in the ACH, wire transfer, and securities services (where scale economies or constant costs are typically found or expected). However, since constant costs now exist for checks (after pricing), this potential for resource misallocation no longer arises.

Another way to view market sensitive pricing is to note that services facing an inelastic demand cross-subsidize other services facing an elastic demand when scale economies exist.[26] As currently applied by the Federal Reserve, such cross-subsidization is limited to individual services within a service line and does *not* occur between service lines. That is, extra revenues are not raised in one service line (e.g., wire transfer) for the purpose of offsetting planned revenue shortfalls in another service line (e.g., ACH). But within a given service line, such as ACH, individual prices are set so that some exceed average costs (debit originator and credit receiver fees) while others are below average costs (debit receiver and credit originator fees). In addition, while total revenues may match total costs (plus the PSAF) for each service line on a national basis, the individual Reserve Banks which supply the nationally priced services (ACH, wire transfer, securities transfer) currently cross-subsidize one another because their cost structures are not identical.

---

[26] The direction of the cross-subsidization should, of course, be reversed when scale diseconomies occur.

Market sensitive pricing, as applied by the Federal Reserve, requires that the price charged for each service be at least high enough to cover the average variable cost of production[27] and, depending on the perceived elasticity of demand for the service, also recover some or overrecover all of the average fixed costs imputed to each specific service. Since the shares of average fixed costs actually covered in prices vary across services within a given service line, some prices will exceed average costs while others are less than average costs. In this manner, the mandated constraint in the MCA that estimated revenues for each service line cover the estimated costs (plus the PSAF) is met. As a result, users who obtain the most benefits from a service, and therefore have the most inelastic demand, pay the most. This same logic lies behind the so-called benefit-flow pricing of the ACH service.

D. Other Pricing Strategies

Several other pricing strategies have been investigated in pricing Federal Rerserve payments services. These are discussed here and include:

(1) Peak-load or time-of-day pricing;
(2) Par value or dollar value pricing;
(3) Benefit-flow pricing; and
(4) Two-part pricing or a fixed charge plus a volume related fee.

Peak-Load Pricing. Also called time-of-day pricing, peak-load pricing is often used in pricing electricity. Its purpose is to lower average costs by inducing a more efficient use of production capacity. Time-of-day pricing was studied for checks, wire transfers, and security transfers but has only been adopted for securities. In all three services, there exist times of the day when a peak demand is experienced which can cause congestion delays in processing checks or funds and securities transfers and induce Reserve Banks to overinvest in processing capacity. Time-of-day pricing would establish prices higher than average costs at times of peak demand and prices lower than average costs for all other times, set so that total revenues match total costs (plus the PSAF) *after* the market adjusts to these prices. The intent is to shift some of the peak volume to an off-peak time where excess capacity likely exists. If peak-load pricing is successful then, over time, the need to invest in new capacity is reduced and existing capacity is more efficiently utilized, thereby lowering costs and prices to users.

---

[27] In practice average variable costs are approximated by the Federal Reserve's allocated direct and support costs for each service listed in the PACS costs accounting publication.

The success of time-of-day pricing depends upon the elasticity of the demand curve for the payments service at the time of peak demand. If the peak demand is very inelastic, then even very high prices will not reduce quantity demanded very much at the peak. If peak demand is elastic, then a high price will reduce peak quantity demanded but, importantly, two things may occur:

(1) Quantity demanded at the peak shifts to the off-peak period for the supplier with peak-load pricing; and/or

(2) Quantity demanded at the peak shifts to the peak or off-peak period for alternative suppliers.

The improvement in capacity utilization is realized only if the first effect is dominant; and the first effect is likely to be dominant only when alternative suppliers do not exist or are of limited importance. This is one reason—but not the most important reason—why time-of-day pricing was adopted for book-entry U.S. government securities transfers in the New York Federal Reserve District (which handles some 70 percent of all such transfers) and rejected for checks and for wire transfers of funds over FedWire in all Districts.

The main reason why time-of-day pricing was rejected for checks and wire transfers is the same reason why its adoption for security transfers has not been very successful in shifting peak volume to off-peak periods (especially after the "moral suasion" effort during its first year of implementation was phased-out). This reason was that the expected reduction in peak demand for checks and wire, or the realized reduction in peak securities demand, was small. A small reduction in peak demand was expected (or realized) because the differential between peak and off-peak prices justified by the capacity cost savings was itself relatively small. While the peak price could easily exceed the off-peak price by 100 to 400 percent, this large percentage differential in prices had a small effect on peak quantity demanded because it was:

(1) Small compared to the average value of the underlying transaction (some $2.5 million per funds transfer or $5 million per securities transfer); and

(2) Small compared to the extra check processing and transportation costs which would be incurred to take advantage of the price differential.

In both cases, quantity demanded either has not been (securities) or was not expected to be (checks, wire transfers) very sensitive to the differentials in peak to off-peak prices which could be defended given the Federal Reserve's own pricing principles. For example, there are only two ways to increase the peak off-peak price differential dramatically to a level high enough to have more than just a marginal effect on reducing peak demand for large value funds and securities transfers over FedWire. One way would be to keep both peak and off-peak prices positive but set the peak price at a level which far exceeds actual cost. Doing this, however, means that total revenue raised would exceed costs for the service line as a whole, raising the possibility that this "excess revenue" would be used to offset losses in other service lines. Such a result would not be consistent with existing pricing principles adopted by the Federal Reserve when pricing was initiated. The only way to have total revenues equal total costs for a service line but still generate a very large differential between peak and off-peak prices would be to make the off-peak price negative. While negative prices could be defended on purely economic grounds, they would not likely be well received by Congress or those FedWire users who continued to use the service at the peak; a negative price makes it more obvious that off-peak users are being cross-subsidized. The situation for checks was different; here a large price differential would have reduced peak demand but, instead of shifting it to the off-peak period at Reserve Banks, more likely it would have shifted to a competitor so that the savings in Reserve Bank capacity costs would not have been realized.

Par Value Pricing. Par or dollar value pricing occurs when the price of the payments service is tied to the value of the underlying transaction rather than being based on its processing cost. Such a pricing arrangement is used in the private sector (e.g., securities safekeeping) and was considered for application to Federal Reserve securities safekeeping and wire and securities transfer services. Par value pricing was ultimately rejected, however, because it was felt that prices are more defensible and better understood by users if they are related to the cost of the transaction rather than to the transaction value.

While this view is still dominant, there exist good reasons to consider an additional charge or fee based indirectly upon the dollar value of a transaction because the risks associated with large-dollar funds and securities transfers over FedWire are directly related to the dollar value being transferred. As discussed in section 8 below, Reserve Banks absorb risk in handling large-dollar transactions and, at present, no risk-related fee is imposed to compensate the Federal Reserve for this "service." In this instance, while an extra charge based on the dollar value of each large-dollar transaction may not be appropriate, it would be possible to levy value-related fees for those large-dollar funds and security transfers which result in so-called daylight overdrafts of reserve accounts, since this is where and when the risk occurs.

Benefit-Flow Pricing. The Federal Reserve has adopted benefit-flow pricing for its wire transfer and ACH services. Benefit-flow pricing attempts to split the price charged between the two parties to a payments transaction—sender and receiver or payor and payee—according to an assessment of the relative benefits obtained by both parties. This is done in such a way that, total revenues cover total costs (plus the PSAF).

In wire transfers, the per transaction charge is split evenly between the party sending funds (debiting the sender's reserve account) and the party receiving funds (crediting the receiver's reserve account). In an ACH transaction, this division generally falls more heavily on the party receiving the funds. Although senders may request that a payment be made, and so typically initiate a transfer, receivers obtain use of the funds received faster than if a check were the means of payment and so also benefit from the transaction. Thus originators of an ACH debit and receivers of an ACH credit gain float compared to a check transaction, which is thought to be a close substitute for an ACH transfer. Originators of an ACH credit and receivers of an ACH debit lose float and thus incur an extra cost compared to a check transaction. These extra costs and benefits have been used, along with the underlying costs of the ACH transactions, to justify the differential prices developed for the ACH service.

While both parties to a check transaction also benefit from it, the Uniform Commercial Code (U.C.C.) governing check payments makes any attempt to assess check collection fees on payor banks very difficult. Indeed, payor banks regularly assess presentment fees on private correspondent collecting banks when checks are presented for payment outside of institutionally established local clearinghouse times. By law, no such charge can be applied to collecting Reserve Banks. The one exception for not charging a payor bank any of the check collection costs may be in establishing a fee for check return items. Although payor banks have the legal right to refuse to pay a check and return it to the collecting bank (acting as agent for the payee), it is the payor's customer who, some 85 percent of the time, causes the check to be returned (usually for insufficient funds in the account). Currently, Federal Reserve return items costs are "bundled" with collection costs and are paid in full by the collecting bank (and the payee). However, this may change in the near future with some (or all) of the return item costs shifted to payors who almost always are the cause of the problem to begin with.

One clear case where the payor bank (and the consumer or corporate payor) benefits from a check transaction is in obtaining check float. While this benefit is large and well-known, Federal

Reserve check prices could not be structured to take this benefit into account for legal reasons, although a strong and clear economic case to support such a price structure can easily be made. In theory, of course, payors and payees negotiate a transaction also taking into account the method of payment—low cost checks with float benefits for the payor vs. higher cost wire transfers with no float benefit. Such negotiation does actually occur for transactions involving large amounts of money. For small dollar transactions, the float cost of checks is absorbed by merchants and, over time, reflected in their prices because of the increased working capital needed to finance the float given to payors. While both parties have in effect shared the check float costs through the "workings of the market," most check users do not face these costs as directly as they would if Federal Reserve check costs (which now include float costs) could be split more equitably between payor and payee banks.

Two-Part Pricing. The final pricing strategy discussed is two-part pricing. This pricing arrangement has been called nonlinear pricing in the economics literature (c.f., Roberts, 1979 and Oi, 1971). It is nonlinear because, if graphed, there is one price component which is fixed—the intercept—and a second component which can vary with volume—the slope, which can be constant or variable. In principle, the fixed part of the price covers the fixed cost of a payments transaction, such as the costs involved in maintaining a wire transfer facility even if no volume is processed—an access fee—or the fixed costs of handling a cash letter (a bundle of checks) regardless of the number of checks in the letter. The second part of the price is a constant or variable fee related to the volume of payments transactions being processed. If this price falls as the volume processed increases above certain levels, it acts as a volume discount. In any case, the second part of the price is set to cover the variable costs of processing a payments transaction.

Two-part pricing is seen as an alternative to Ramsey pricing with a revenue constraint. Under two-part pricing the volume-related fee will be set at or close to marginal costs so the price of using additional amounts of the service is close to the real resource costs involved, just like marginal cost pricing. The fixed fee is used to recover fixed costs so that, unlike pure marginal cost pricing, a revenue constraint can be met.

One problem with two-part pricing is that some users may be induced to act as an intermediary and resell the service to others after paying the fixed fee. The volume-related fee of course is unchanged in the reselling process but the fixed fee can be spread over all users who repurchase the service through one user rather than all users paying a separate fixed fee to the original supplier. Since the fixed fee is initially set to recover fixed costs from a predetermined number of users, a reduction in the number of users paying the fixed fee to the original supplier can result in a revenue shortfall for the supplier. At the limit, if reselling can not be limited or prohibited, two-part pricing runs into the same revenue problem faced by marginal cost pricing. While this does not yet appear to be a problem for the Federal Reserve in services where two-part pricing has been adopted, it could easily come about if the fixed fee is set high enough to induce such behavior, even if fixed costs are in fact larger than the fixed fee charged.

E. Similarities Between Correspondent Bank and Federal Reserve Pricing Strategies

There are some important similarities between correspondent bank and Reserve Bank pricing of payment services. Both consider, in a nonrigorous way, the likely own and cross-price elasticities of demand in setting prices for a particular service within a service line (e.g., the price of one type of check service within the check service line). This is a direct application of Ramsey pricing, called value pricing in some of the commercial bank pricing literature (Guiltiman, 1980 and American Bankers Association, 1982). Value to a user is directly reflected in his price elasticity; the more value a service has, the more inelastic the demand is likely to be so prices would tend to exceed average costs here. The reverse holds for those payment services where perceived value is relatively low.

While both groups in effect practice Ramsey pricing within a service line, Reserve Banks do not practice it *between* service lines. That is, revenues higher than costs in one area (like wire transfers) are not used to cross-subsidize revenues lower than costs in another area (like checks). Such a restriction was not in the MCA but is reflected in certain "pricing principles" later adopted by the Federal Reserve. No such restriction, of course, applies to correspondent banks which, at least in the past, apparently commonly engaged in cross-subsidization across service lines. However, large banks which have more recently set up their payment service lines as separate profit centers have chosen to forego Ramsey pricing between service lines in the belief that overall cost-efficiency is better attained when such cross-subsidization between service lines is eliminated or at least minimized.

V. The Private Sector "Markup" and Pricing Federal Reserve Float

Two other Federal Reserve payments pricing issues concern the so-called private sector markup and the pricing of float. Both are important in determining the relative costs of different payment instruments.

A. Pricing Federal Reserve Services with a Private Sector "Markup"

While Federal Reserve payments services are primarily determined using market sensitive pricing, they are set to generate revenues sufficient to recover all Federal Reserve costs plus a private sector adjustment factor or PSAF. A PSAF is necessary since the Federal Reserve pays no Federal taxes, provides only a statutory determined return—6 percent—on member bank held equity, and does not actually finance the acquisition of physical assets. In practice, such assets are fully paid for in the year acquired by returning less surplus funds to the Treasury. The MCA requires that Federal Reserve prices reflect a PSAF comprised of:

> ". . . an allocation of imputed costs which takes into account the taxes that would have been paid and the return on capital that would have been provided had the services been provided by a private business firm [MCA, Section 107]."

The problem is that there is no private sector firm which only provides all the payments services provided by Reserve Banks. Thus what appears to be a straight-forward accounting exercise is one which has been fraught with many theoretical and practical problems in application because no easily identifiable private sector counterpart is available.

Ideally, the payments processing portion of a large commercial bank's correspondent banking operations should be used. But such separate information is typically not publicly available. Since the most desirable information source is unavailable, the Federal Reserve had to choose among a number of less desirable alternatives, such as:

(1) A set of large bank holding companies who provide correspondent bank payment services as *one part* of their service output;

(2) A sample of nonbank data processing companies who provide

some types of correspondent banking payment processing and accounting services for banks; or

(3) A sample of regulated public utilities or government-sponsored entities (such as the Federal National Mortgage Association).

The third alternative was viewed as inappropriate since tax rates, debt/equity ratios, and the cost of equity for public utilities or government sponsored entities was felt to be unduly influenced by the regulated status (for public utilities) or by the ability to access funds at less than market determined rates (for government-sponsored entities).

The Federal Reserve chose to base its PSAF on a set of large bank holding companies (BHCs), adjusting for their exceptionally low tax rates and their high leverage ratios. Of the three essential elements of the PSAF:

(1) Tax rates;
(2) Capital structure (debt/equity ratio); and
(3) Return on capital (the weighted average of debt costs and return on equity);

the Federal Reserve adjusted the first two to account for the noncorrespondent business of banks and took the third basically as published in BHC annual reports.[28]

The statutory total marginal federal income and state tax rate applicable to BHCs is currently 51 percent but, largely because of nontaxable income from state and municipal security holdings, the actual marginal tax liability is around 42 percent. When account is taken for those taxes which are in fact paid, leaving out those which are deferred to a future date, the BHC tax rate falls to around 20 percent. But when this paid rate (which excludes deferred taxes) is adjusted for bank holdings of tax free securities (and extraordinary gains and losses on all security holdings), the rate rises to around 36 percent. This is the rate which has been proposed by the Federal Reserve for 1984; it reflects taxes paid, adjusted for tax-free income which a "stand alone" correspondent banking operation would be unlikely to receive.

Capital structure at large BHCs is basically determined by the deposit taking and intermediation functions inherent to banking. Because of this, equity capital rarely exceeds 5 to 6 percent of total assets at large banks. Rather than guess at what the debt/equity ratio might be in a "stand alone" correspondent operation, the Federal Reserve determined its capital structure largely by matching the life or type of priced service asset on its books with the most common method of financing these assets in the private sector. Total Reserve Bank priced service assets were first decomposed into long-term and short-term assets. Then the long-term component was assumed to be financed at the same ratio of long-term debt to equity as that observed at a set of large BHCs. For example, Reserve Bank premises comprised some 62 percent of total priced service assets to be financed and so is matched with a capital structure in which equity makes up some 64 percent of the total. Priced service equipment and furniture was assumed to be financed by long-term debt (27 percent) and short-term assets (receivables and supplies) were financed through short-term debt (9 percent).[29]

---

[28] This is an oversimplification because the rate of return on equity is affected by decisions on the tax rate and the cost of debt used excluded the interest cost on deposits with regulatory imposed interest ceilings.

For 1984, the main components of the proposed PSAF were:

(1) Tax rate: 36 percent;

(2) Capital structure: short-term debt (9 percent), long-term debt (27 pecent); and equity (64 percent); and

(3) Return on capital: short-term debt rate (9.5 percent), long-term debt rate (10 percent), and pre-tax return on equity (21 percent).

Putting all this together gives a weighted average cost of capital of some 17 percent. Applying this cost of capital to the priced service assets to be financed adds $56 million to the direct, support, and overhead costs of some $389 million to be recovered in 1984.[30]

It is at this point that a great deal of public confusion occurs. Taking the $56 million in financing costs to be recovered in the PSAF as a percent of total Reserve Bank expenses to be marked up of $389 million gives a "mark-up" of 14.5 percent. In public discussions, this mark-up has been confused with a return on equity or a mark-up on direct costs. As seen, the return on equity (pre-tax) is 21 percent, not 14.5 percent. And if the traditional mark-up on direct costs were used for Reserve Banks such a mark-up would be 69 percent since both the $56 million in financing costs plus the $125 million in Reserve Bank overhead would be captured in the traditional private sector mark-up over direct and support costs of $264 million. Transportation expenses of $79 million are not marked up (but are recovered in the price). They are not marked up because they include almost no Reserve Bank costs, reflecting as they do private sector contracts to provide transportation services to users of Reserve Bank payments services. As such, they already include a market rate of return or profit.

---

[29] Clearing balances represent a "self-financed" asset since these balances are held idle by Reserve Banks, earn a market return, and pay out a market return which is then used to pay for Reserve Bank payment services used by the (bank) owner of the clearing balance. Clearing balances are thus used to pay for Reserve Bank services and are a perfect substitute for fee income; in no way are they necessary to the *production* of payment services as such, although they are also used like reserve accounts to clear and settle payments. This parallels bank use of correspondent balances to pay for and clear respondent bank payment services.

[30] The 1984 data on the PSAF were obtained from an October 13, 1983 *Federal Reserve Press Release* on this topic.

It has been argued by Koot and Walker (1980) and by an interested coalition of large correspondent banks that the Federal Reserve's choice of a BHC model for computing the PSAF is misplaced. Nonbank data processing companies should be used instead. But just as large commercial banks perform *more* services than just those payment related services provided by Reserve Banks, nonbank data processing companies perform *fewer* services than Reserve Banks. In particular, while providing check processing services, they do not collect checks from payor banks. Nor do they currently operate ACH facilities, offer large dollar funds and securities transfers, provide security safekeeping services, or sort and ship currency and coin. Just the same, adjustments could be developed to compensate for these deficiencies just as they were developed to compensate for the fact that large commercial banks provide more services than just payments services.

The real point at issue, however, goes beyond the appropriate model to use for the PSAF. The underlying issue is simply that large commercial banks *are* the competitors of Reserve Banks and, because they offer a more diverse range of services than Reserve Banks, can cross-subsidize their correspondent payment services operation by market sensitive pricing of their deposit, loan, and other services. Put differently, if payment services at large banks do not "stand alone" as separate profit centers with *separate financing of assets*, neither should they at Reserve Banks.[31] To do so would place restrictions on Reserve Banks not also placed on their competitors—the large commercial banks.

The cross-subsidization issue is not unique to commercial bank-Reserve Bank competition in providing payments services. It has arisen between I.B.M. and other (more specialized) computer hardware suppliers and between A.T.&T. and suppliers of alternative long-distance communication facilities. In each case, the supplier of a more narrow range of services or a supplier of a broad range of services to users with elastic demands, argues that "fair" competition is difficult when:

(1) The competitor supplies a broader range of services and thereby possibly enjoys economies of scope (where joint costs are spread over a larger range of services resulting in lower prices); and/or

---

[31] Treating payment services as a separate profit center within a large commercial bank is a necessary but not a sufficient condition for "fair" competition to exist. Separate financing of assets at a different debt/equity ratio, without deposits counted as debt, is also necessary.

(2) The competitor supplies the same broad range of services but, because different markets are served, faces users with more inelastic demands (on average) so that market sensitive pricing permits greater cross-subsidization of services.

Of course, there is no real solution to this problem other than to let market forces work or make a strong enough case that the cross-subsidization which occurs is not in the public interest (c.f., the A.T.&T. court ordered breakup which separated local from long distance communication).

B. The Pricing of Federal Reserve Float

The Monetary Control Act (MCA) requires the Federal Reserve to charge for the float it creates at the federal funds rate. Federal Reserve float is created when credit is passed to check and ACH depositors, securities sellers, and wire transfer receivers before Reserve Banks obtain credit themselves.

Prior to pricing, Federal Reserve float was, in effect, an interest-free loan to the banking system. As seen in table 6, this loan had increased from $2.7 billion per day in 1973 to $6.5 billion in 1979. Evaluated at the average federal funds rate of 11.2 percent in 1979, banks and their customers received some $728 million in before-tax income from Federal Reserve float. With the passage of the MCA in 1980, float has fallen to $1.7 billion in 1983, the year it was priced and incorporated (for checks) into the price schedule. Prior to 1983, float was not priced but was instead reduced through operational changes and expedited check collection procedures. When float is priced and included in the cost base to be recovered from depositing banks, these costs will vary as the federal funds rate changes. That is, changes in the federal funds rate will shift the average and marginal cost curves shown in figure 2 above. Thus the price to depositing banks is higher the more that float is recovered by including it in the cost base (rather than reducing float through operational improvements) and has greater variance than if the cost base only included operating costs (which are more stable than float evaluated at the federal funds rate).

Reducing Float Versus Pricing Float. Pricing float at the federal funds rate or reducing it has the same effect on Treasury revenues (which rise). Since reducing float through operational changes was initially much more cost effective to the users of Reserve Bank payment services, this approach was taken first until these opportunities were basically exhausted. At that point the float remaining was evaluated at the federal funds rate and directly added to the cost of the priced service. When float was being reduced through operational improvements, the initial cost effectiveness was quite high. During 1980 and 1981, for example, each $1 in extra operational expenses was expected to be associated with a $5 reduction in the actual cost of float (and a $450 million reduction in the amount of float outstanding).

An important additional consideration behind the Federal Reserve's decision to pursue operational improvements first and direct pricing of float second concerns equity. In the case of Federal Reserve check float, which in 1979 comprised more than 90 percent of total FR float, payors or check writers directly gain from float. The total float gain includes mail float, bank processing float, as well as FR float. In choosing to write a check rather than send cash or a wire transfer (where payment to the receiver is in immediate or same day funds), the check writer makes payment in funds with delayed availability. Since check writers experience a float benefit from using checks, equity would argue that payors, not payees, should bear the full cost of recouping the Treasury revenues lost from FR float. Following this reasoning, serious thought was given to directly charging payors for FR float. However, this was ultimately rejected for legal reasons. Given the focus of the Uniform Commercial Code—the code which governs the rights and liabilities of check payors and payees—it was determined that charging the paying bank (payor) for FR float was less defensible than charging the depositing bank (payee).[32] Equity was still served, however, by instituting operational improvements which reduced the level, and hence the benefits, of FR float to paying banks and their customers. Although the current (lower) level of check float is now priced and paid for by depositing banks, the main brunt of the costs of FR float—some 75 percent of the total—was and continues to be borne by paying banks because of operating improvements which speeded up collection to payors and reduced their benefit from FR float.[33]

The Real Resource Cost of Pricing or Reducing a Transfer Payment. From a pure efficiency or resource allocation standpoint, economists often argue that expenditures of real resources to reduce float are socially wasteful. This view correctly notes that float is a transfer payment from payees to payors and that society as a whole—that is, payees *plus* payors—are made worse off when one group or the other spends money to alter the level or distribution of float between them. The real resources spent to slow down (remote disbursement) or speed up (expedited collection) check processing, collection, and settlement could instead be spent on producing more real output, rather than redistributing that which already is produced. This argument, while economically correct, neglects the following points:

---

[32] One application where the legal case for charging the payor bank for float is viewed to be more defensible is presented below when Electronic Check Collection is discussed.

[33] This figure is derived by taking the reduction in FR float between 1979 ($6.5 billion) and 1983 ($1.7 billion) and expressing it as a percent of the float level in 1979.

(1) Revenue loss—Congress was presented with the transfer payment argument prior to the passage of the MCA (c.f., Wolkowtiz and Lloyd-Davies, 1979) and rejected it, choosing instead to reduce the revenue loss to the Treasury from FR float;

(2) Distributional equity—even if Congress funded FR float out of general tax revenues, the incidence of these taxes is likely to be quite different from the distribution of float benefits to check users. Federal taxes fall most heavily on middle income individuals while FR float benefits are concentrated at those businesses which write a large volume and dollar amount of checks; and

(3) Negotiation difficulties—with large numbers of payors and payees it has not been possible for these two groups to negotiate the distribution of any meaningful amount of the check float benefits and permit more cost effective payment arrangements to be used, saving real resources in the process.

Only when few payors and few payees are involved, and when the real resource cost savings can be captured by these same parties, has there been successful negotiation of the benefits of check float, permitting more cost effective payments procedures to be utilized. This has occurred in the ACH Corporate Trade Payments Program where the 10 pilot participants[34] have:

(a) Calculated the average float obtained from check disbursements between themselves;

(b) Agreed to alter their trade credit terms to one another to offset the lost float benefits of using ACH transfers in place of checks; and

(c) Saved real resources by automating their internal processing of accounts payable and receivable by placing accounting, invoice, and posting information in the addenda records to ACH payment files.[35]

While ACH processing of payment information can be cheaper than writing, disbursing, and otherwise handling checks, this is not where significant real resources are saved. The real savings come from automating the other related payment/accounting/posting operations associated with the complete processing of payment information.

---

[34] The Corporate Trade Payments participants, among others, include Sears, Westinghouse, Black & Decker, and Exxon.

[35] Normal ACH payment information contains only the identification of the paying and receiving financial institution along with the date and amount of the payment. In the Corporate Trade Payments Program additional information on the corporate paying and receiving institutions are added, such as the amount and number of different invoices for which total payment is being made and other information regarding trade credit, late delivery, and returned goods which affect the payment value.

Pricing of Float. The distribution of Federal Reserve check float in 1982 is shown in table 6. Inter and intraterritory float involves transportation related check float, between and within FR office operating territories. Transportation float is the result of unfavorable weather conditions and other unexpected transportation delays in presenting checks for payment after credit is passed to depositors according to preestablished funds availability schedules. This is the float that is priced to each of the over 600 *individual* depositing banks which directly deposit certain types of interdistrict items. Holdover, intraterritory, and return item float, in contrast, is now evaluated at the federal funds rate and incorporated directly into the check processing prices paid by *all* depositing banks (except for large dollar interterritory return items where float is avoided by deferring credit). This different treatment occurs because holdover and return item float are more stable and dependent on the structure of internal operations at

TABLE 6

Federal Reserve Float: Level and Distribution
(1973-1983)

Federal Reserve Float Levels (annual average of daily levels; $ billions)		Federal Reserve Float Distribution (1982; $ billions)		
1983	$1.7	Check Float		$1.76
1982	2.3			
1981	3.2	1. Interterritory	$1.00	
1980	4.5 (MCA passed)	2. Holiday and Midweek Closing	.26	
1979	6.5			
1978	5.4	3. Holdover	.20	
1977	3.8	4. Return Item	.15	
1976	2.6	5. Intraterritory & Other Check	.15	
1975	2.1	Noncheck Float		$ .54
1974	2.3	Total		$2.30
1973	2.7			

Sources: *Federal Reserve Bulletin*, various issues; and Pricing Policy Committee, "Federal Reserve Float Reduction Proposals," memorandum to Board of Governors of the Federal Reserve System, February 14, 1983.

Reserve Banks (and thus absorbed in the price schedule or eliminated by adjusting availability schedules) while transportation related float is less predictable and more external to Reserve Bank control (and thus passed on directly to those individual depositing institutions which obtain credit before collected funds are received by Reserve Banks).[36]

Depositing banks have a number of alternatives in paying for inter-territory or transportation-related float, the main component of current check float. These funds availability and payment choices are:

Crediting Option	Payment Method
Fractional (partial day) availability (availability known beforehand)	(no float created *on average*; no charge)
Fixed (whole day) availability (availability known beforehand)	Float created is paid through an explicit charge, an increase in clearing balances, or by an "as of" adjustment to reserve balances.

The fractional availability choice creates no float on average over time so no float payment is neccessary. Another "no charge" method initially proposed was actual availability. However, the passing of credit with actual availability is uncertain and depends on when the deposited checks are actually collected. As there was little interest in this option, it was dropped. The fractional and fixed availability options do not have this uncertainty and were retained. With fixed funds availability, the crediting arrangement is in whole day increments which are known beforehand, such as all credit on the same day or the next day after checks are deposited at Reserve Banks. With fractional availability the same day fraction could be .90 (not 100% as under fixed availability), with the remaining .10 being credited the next day. The fractions used are based on special samples of the actual time it takes to collect check deposits and are altered over time to reflect historical experience. Since over 85 percent of the 620 depositing banks who can choose a crediting option, chose the fixed availability option, most institutions would not be affected if the fractional availability option were dropped in the future.

---

[36] Holiday and midweek closing float occurs when Reserve Banks accept deposits and pass credit but are themselves unable to collect the items. This is because banks in another FR District are closed due to a holiday or because it is their custom (in certain midwestern states) to be closed on a weekday but open on Saturday (when Reserve Banks are closed). A final policy on how and who to charge for this type of float is not yet in place.

The payment methods for fixed availability involve either exactly offsetting holdings of balances or the payment of direct fees (after evaluating the float created at the current federal funds rate). The extra idle credit earning clearing balances held or the "as of" adjustments to reserve balances act as compensating balances for Reserve Banks. This replicates the popular charging method used by commercial banks.

Monetary Policy Effects. Federal Reserve float augments the money supply since loanable funds are being supplied to the banking system. Under a reserves targeting procedure, it is neccessary to forecast FR float so that this impact on the money supply can be predicted. When funds availability is known beforehand, as with the fractional and fixed availability options, banks know their funds position but the Federal Reserve has to forecast the daily level of float created. As the average daily level of float has fallen due to the implementation of operational improvements, the day-to-day variance of the float which has to be forecasted has also fallen. As a result, there has been some minor improvement in the Open Market Desk operations and information flow due to the reduction and pricing of FR float.

VI. Scale Economies in Providing Payments Services

Scale economies are important because they influence the choice of a proper pricing strategy by suppliers. Certain pricing strategies, as noted in section IV, are optimal when scale economies can be shown to exist but are nonoptimal with scale diseconomies. In addition, by measuring the scale economies of different types of payments services we can determine whether or not resources are being efficiently allocated between them. From this it is possible to determine the need for governmental intervention limiting new entry or, alternatively, increased private sector competition to improve resource use.

Scale economies exist when unit or average costs fall as output is expanded. Cost and production functions are used to estimate statistically how processing costs vary with the scale of operation of a Reserve Bank or a commercial bank. These statistical models are at times quite complex and the following discussion has been simplified, with most of the technical details deleted, relegated to footnotes, or placed in an appendix. In what follows, tabular analysis and statistical models have been used to illustrate scale economies in two distinct payment service areas:

(1) Wholesale payment operations, such as Reserve Bank check, ACH, and wire transfer processing (with references to other studies dealing with book-entry securities transfers and coin and currency operations); and

(2) Retail payment operations, such as commercial bank provision of deposit and loan services.

While some of these payment services experience strong scale economies, others do not. The implications of these results for future cost/price induced substitution between different payment methods—the payments system's product mix—are discussed below in section IX.

A. Wholesale Payments Operations: Scale Economies in Federal Reserve Payments Services

Use of Cross-Section Data. Scale economies (diseconomies) exist when long-run average costs fall (rise) as check, ACH, or wire transfer processed volume rises. The behavior of long-run costs at different levels of processed volume or output is assumed to be adequately reflected by estimating statistically the association between FR office annual check, ACH, or wire transfer total activity or production costs and annual volume levels in a cross-section of from 36 to 48 offices, holding other influences on costs constant. This cross-section approach is commonly used in the literature because it minimizes two important problems:

1. The difficulty of obtaining consistently defined cost accounting data over long enough time periods to perform time-series analyses; and
2. The difficulty of accurately measuring technological change and other time-related cost influences.

The basic assumption underlying the use of cross-section data to measure scale economies is that those offices which now process small volumes would resemble the larger offices in the cross-section if their scale of operation were to be expanded and managers were given time to make the proper adjustments in the size and scope of their operations. The adjustments would include an optimal amount of excess processing capacity to meet peak volume levels without generating an excessive amount of holdover float. Thus the statistical model concerns long-run equilibrium cost adjustments to larger processed volume levels; we are not dealing with the costs associated with short-run fluctuations in output.

A Cost Function Model. Scale economies should be estimated using an approach which has (at least) the following two properties:

1. The model used should be flexible enough to reflect scale economies, constant costs, and scale diseconomies if these three ranges to the average cost curve exist simultaneously in the data; and
2. The scale economy results should be evaluated at several different sized operations in addition to evaluation at the average size of all operations.

The first property rules out the use of certain models used by others, such as the Cobb-Douglas model which has been often used to estimate scale economies in commercial banking.[37] What is required is a quadratic functional form like the translog cost function model specified in the appendix, to which the interested reader is now referred for a more detailed exposition. This statistical model is consistent with earlier analyses of check, ACH, and wire transfer scale economies (c.f., Humphrey, 1981 and 1982). The second property permits us to see how scale economies might vary as check, ACH, or wire transfer volume moves from small to larger levels and provides more information than if one looked only at the results for the average volume office alone. Statistical tests can then be performed to determine whether a significant difference exists by size of operation, permitting the model to be simplified if this hypothesis is not rejected.

---

[37] One well-known and comprehensive study which has earlier used this approach is Bell and Murphy (1968). A more recent example is Clark (1984).

B. Estimates of Federal Reserve Scale Economies

A Look at the Raw Data. Before presenting the statistical measures of scale economies, it is instructive to look at the raw data to determine how check, ACH, and wire transfer average costs vary across four size-classes of Federal Reserve offices. These data are shown in table 7. Average unit costs for checks (1983 data) are seen to describe a U-shaped cost curve as processed volume moves from the lowest to higher volume offices.[38] But while check processing appears to face a U-shaped cost curve, ACH average unit costs (1982 data) steadily fall across the four size-classes. The wire transfer data (1983), in contrast, appears to have a U-shaped relationship. Looking only at the raw data, one would conclude that ACH experiences only scale economies while check processing experiences scale economies for the first three size-classes but slight scale diseconomies at the largest size-class. Wire transfers appear to experience slight scale economies for the first two size classes with strong diseconomies for the last one.

It is possible to compute an overall implied scale economy measure from these raw data, shown in the last row of table 7. Such a measure is theoretically imperfect, since non-scale influences on costs are not held constant, but it at least captures the overall change in total activity costs and processed volume between the smallest and largest size-classes. Scale economies (SCE) exist when SCE = (% change in total costs)/(% change in output) is less than 1.0 since total costs rise less than output and average or unit costs fall as more is produced. Scale diseconomies occur when SCE is greater than 1.0 since total costs rise more than output and average cost increase. Constant costs exist when SCE = 1.0 and the average cost curve is flat or constant.

For ACH, since this scale measure is very much less than 1.0, strong scale economies are implied, although this implied measure far overstates their importance when compared to the statistically derived results presented below. For checks, slight scale diseconomies are implied since this measure is greater than 1.0. Wire transfers appear to evidence strong scale diseconomies but, since the last size

---

[38] The check offices which comprise these four size-classes are shown below in table 8. The ACH and wire transfer offices in each size-class are not separately shown, although there was a considerable degree of overlap among all three sets of offices within each of the four size-classes.

TABLE 7

Federal Reserve Check, ACH, and Wire Transfer Unit Costs by Four Size-Classes of FR Offices (average costs; $ per 1,000 items)

FR Office Size-Class		Checks (1983)	ACH (1982)	Wire (1983)	Approximate Percent of All Office Volume
I	(Smallest Offices)	$14.70	$64.30	$467.25	16%
II		13.70	51.70	447.17	20%
III		12.90	46.50	465.84	27%
IV	(Largest Offices)	15.00	40.80	575.96	37%
All Offices		14.10	48.29	501.56	100%
Implied Scale Economies from size-class I to IV		1.04	.35	1.42	

Source: Computed from Board of Governors of the Federal Reserve System, *Planning and Control System Expense Report*, detailed annual cost accounting reports, 1982 and 1983.

class in fact contains only two FR offices, this result is not supported in the statistical model where the hypothesis of constant costs can not be rejected statistically. That is, constant costs dominate.

Check Processing Scale Economies. Check scale economy estimates from the statistical model presented in the appendix are shown in table 8. Because the underlying statistical model for checks includes a quadratic term, the scale economy measure (equation 4 in the appendix) can vary as check volume is altered. Thus each of the 48 check offices shown in table 8 has its own scale economy measure. The offices are ranked according to their post pricing 1983 check volume level (column 5), lowest volume offices first. The effects of pricing on check volume growth (not shown) have been quite diverse, with some offices experiencing decreases and others increases over the seven year period 1976-1983 under study.[39]

The statistically estimated check scale economy results prior to pricing (1976) are shown in columns 1 and 2. Those shown in column 1 are from a very sophisticated modeling and estimation effort

---

[39] Frodin (1984) contains a more detailed description of this result by focusing upon the emergence of new or renewed local check clearinghouses which have been activated since pricing.

TABLE 8

Federal Reserve Check Scale Economies Over 1976-1983

(SCE = percent change in total production costs/percent change in output)
(Q = total processed commercial checks in thousands; AC = average production costs for processed checks)

	Office	(1) Original 1976 SCE	(2) SCE 1976	(3) SCE 1982	(4) SCE 1983	(5) Q 1983	(6) AC 1983
I.	El Paso	.73*	.75*	.68°	.72	45,636	21.3
	Lewiston	a	a	.74°	.76	58,490	14.9
	Helena	.76*	.77*	.76°	.80	72,614	16.3
	Louisville	.94	.91	.81°	.82	84,327	22.9
	Memphis	.93	.89°	.84°	.82	85,057	21.9
	Little Rock	.90	.86°	.83°	.85	100,635	16.8
	Charleston	d	d	.84°	.87	108,934	17.9
	Columbia	e	e	.87	.87	111,592	18.5
	Omaha	1.00	.94	.91	.89	122,717	19.5
	Portland	.92	.92	.90	.90	132,371	15.2
	San Antonio	.92	.90	.90	.90	132,961	18.6
	Salt Lake City	.91	.89°	.89	.91	138,072	15.0
	Pittsburgh	1.02	.98	.95	.93	164,313	21.4
	Columbus	c	c	.92	.94	167,439	13.9
	Buffalo	.87	.86°	.93	.94	171,287	15.8
	Milwaukee	f	f	.97	.95	178,752	19.0
	Birmingham	1.02	.96	.97	.97	197,644	13.3
II.	Houston	1.05	.98	.98	.97	201,220	17.1
	Cincinnati	1.06	1.00	1.00	.97	203,078	17.8
	Oklahoma City	.99	.94	.98	.97	204,805	14.5
	Seattle	1.01	.96	.98	.98	211,089	11.2
	Utica	b	b	1.00	.98	219,504	18.1
	Indianapolis	f	f	.98	.98	220,447	12.7
	Nashville	1.02	.95	.98	.99	223,318	12.8
	Kansas City	1.17*	1.07°	1.02	.99	229,758	22.0
	Des Moines	f	f	1.02	1.01	250,930	15.7
	Windsor Locks	a	a	1.06	1.02	269,829	16.5
	St. Louis	1.17*	1.07°	1.05	1.02	281,097	18.9

TABLE 8 (continued)

Office	(1) Original 1976 SCE	(2) SCE 1976	(3) SCE 1982	(4) SCE 1983	(5) Q 1983	(6) AC 1983
III. Baltimore	1.18*	1.08°	1.05	1.03	282,729	22.7
Richmond	1.12*	1.09*	1.04	1.03	284,018	17.1
Charlotte	1.19*	1.04	1.05	1.03	286,109	16.6
Cleveland	1.18*	1.09°	1.05	1.03	291,659	17.0
New Orleans	1.06	1.00	1.04	1.03	295,230	12.7
San Francisco	1.07	1.02	1.07	1.04	315,842	19.0
Jericho	b	b	1.07	1.05	332,444	15.0
Detroit	1.13*	1.05	1.08	1.06	339,129	15.3
Cranford	b	b	1.09	1.07	378,460	15.4
Atlanta	1.15*	1.06	1.11	1.08	396,441	14.0
Miami	g	g	1.11	1.09	409,668	12.7
IV. New York	1.39*b	1.27*b	1.14	1.09	424,312	34.7
Philadelphia	1.25*	1.13*	1.12	1.09	425,226	19.7
Los Angeles	1.12*	1.08°	1.12	1.09	429,404	14.7
Denver	1.15*	1.07	1.13	1.10	445,355	14.8
Boston	1.33*a	1.21*a	1.15	1.10	449,747	20.1
Jacksonville	1.31*	1.17*	1.15	1.11	457,965	11.4
Dallas	1.19*	1.08°	1.17	1.13	529.786	17.3
Minneapolis	1.25*	1.13*	1.24	1.20	782,042	17.4
Chicago	1.46*f	1.28*f	1.26	1.20	820,347	20.8

Source: Column 1 is from Humphrey (1981); Columns 2 to 4 are estimated from the statistical model; Columns 5 and 6 are from *Planning and Control System* (PACS), annual cost accounting reports concerning check processing output and total activity costs. Total activity cost excludes certain overhead costs but covers the activity or production costs for processed checks, return items, and adjustments. Fine sort production costs were less than 4 percent of this and were excluded, along with fine sorted output which was not included in Q.

Notes:

a—Lewiston and Windsor Locks have been merged with Boston.

b—Jericho, Cranford, and Utica have been merged with New York (Utica started in November 1976).

c—Columbus has been merged with Cleveland.

d—Charleston has been merged with Richmond (Charleston started in November 1976).

e—Columbia has been merged with Charlotte.

f—Milwaukee, Indianapolis, and Des Moines have been merged with Chicago.

g—Miami has been merged with Jacksonville.

*Indicates that the SCE estimate is significantly different from 1.0 or constant average costs at the 95 percent confidence level using a two-tailed *t* test;° shows the 90 percent confidence interval.

reported earlier (Humphrey 1981).[40] Those in column 2 were based on the model shown in the appendix, which used data constructed in the same manner (except for the price of capital) as in the earlier study. As is seen, these two sets of results are very similar numerically, illustrating the robustness of the relationship being measured. Scale economy values which are starred (*) or circled (o) indicate results which are significantly greater than (showing diseconomies) or less than (showing economies) 1.0, which is the constant cost case. In 1976, prior to pricing, 13 offices (column 2) to 17 offices (column 1), out of 36 offices, are shown to have experienced statistically significant scale *diseconomies*. After pricing in 1981, the number of such offices declined to zero in 1982 (column 3) and 1983 (column 4), out of 48 offices. Put differently, prior to pricing from 67 to 78 percent of all Federal Reserve deposited checks were processed at offices which experienced significant scale diseconomies. This inefficiency of resource use has declined dramatically after pricing in 1981. In 1982 and 1983, no checks were processed at offices experiencing significant scale diseconomies. Only when the confidence level is lowered to 80 percent are some significant check scale diseconomies observed, and then only in 1982, not 1983. On balance, the pricing of the Federal Reserver's check service has clearly improved resource allocation for society as a whole. After pricing, almost all FR offices experienced constant costs in their check processing operations.

There are at least two good reasons why the observed improvement in resource allocation has occurred:

(1) The Reserve Bank check service was priced; and
(2) A requirement to match revenues with costs (including the PSAF and, most recently, Federal Reserve check float) was imposed.

Services provided at a zero price can be overused since banks will purchase Federal Reserve check services up to the point where their explicit cost, which was zero, equals the value the extra service used has for the bank.[41] This apparently led to the scale diseconomies observed in check processing prior to pricing. Check volume, with few exceptions, was demand determined; there was no price incentive to limit its use even if costs rose as volume expanded.[42] And with a zero price there was no signaling mechanism available to induce new suppliers to offer (at a profit) a competitive service of processing low dollar amount checks for many endpoints using favorable Federal Reserve funds availability schedules.[43] With Federal Reserve pricing, users purchase less of the service, overuse declines and, importantly, the private sector is now able to enter the low dollar amount check processing market if it can provide the service at a lower price (assumed to reflect lower costs as well, although this may not always be the case). In a competitive market with reasonably free entry and use of similar production technology, the end result for all suppliers of check services should be that they are producing on the flat portion of their average cost curve. That is, the Federal Reserve scale diseconomies observed prior to pricing will be reduced or eliminated once pricing occurs within a competitive market for check services and constant costs should be the rule. This explanation is consistent with our empirical results over the seven year period studied, as seen in table 8. While some of the post pricing scale economy values shown in table 8 are numerically greater than 1.00, similar to the calculation using raw data in table 7, these values were not significantly different from the constant cost case of 1.00. Hence constant costs were the rule after pricing.

---

[40] In this earlier study, in order to make consistent comparisons to years prior to 1976 and because two offices started operations in late 1976, eleven new regional check processing centers (Lewiston, Windsor Locks, Jericho, Cranford, Utica, Columbus, Charleston, Columbia, Milwaukee, Indianapolis, and Des Moines) and one new branch office (Miami) were combined with offices which had just prior to or during 1976 been responsible for processing the checks now processed by these twelve offices. This aggregation is also performed in column 2 in order to more accurately compare the simplified model's results to those reported earlier for 1976. The 1982 and 1983 data are not aggregated so 48 offices are shown, rather than the 36 offices used in 1976.

[41] While Federal Reserve earnings on idle reserve balances more than covered the cost of Reserve Bank payment services, these balances were not explicitly tied to use of these services. Reserve requirements thus functioned as a "fixed fee" for access to services, at which time any amount could be used without charging an additional fee. In this sense, Reserve Bank payment services had a zero price for using more or less of the service once access was achieved.

A second reason why check diseconomies have been reduced probably lies in the MCA requirement that revenues match costs, including the PSAF and Federal Reserve Float. The cost-revenue match requirement, due to its ease of verification and because of its relatively strict implementation, has likely served as a more effective management tool to control costs than the internal reviews and informal guidelines relied upon prior to pricing. For these reasons—pricing, a competitive market, and improved cost control incentives—resource allocation in check processing operations has on balance been improved by the passage of the MCA.

ACH Scale Economies. As noted in the appendix, the statistical model used to estimate ACH scale economies does not require a quadratic term; these scale economies are the same for each office in a given year but differ slightly across the five years studied (1978-1982). It is of interest to see if these scale economies differ as well according to their head office or branch location. Thus ACH scale economies are presented each year separately for:

---

[42] Check volume was to some degree regulated through nonprice restrictions, such as the restriction against accepting deposits from banks which were drawn upon local institutions belonging to private clearinghouses in the same city as a Federal Reserve office.

[43] For large dollar amount checks, the private sector was able to effectively compete because the float gain from expedited private sector check processing and collection more than offset the positive price that was charged for this service by correspondent banks. At that time, Reserve Banks did not outsort large dollar amount checks for expedited collection and thus, in effect, chose not to compete for this segment of the market.

- All 38 ACH offices together;
- Only the 12 ACH offices located at Reserve Bank head offices; and
- Only the 26 ACH branch and Regional Check Processing Center offices.

The ACH scale economies estimated for each of the five years before and after pricing in 1981 are shown in table 9. Over the five year period as a whole, a 10 percent increase in total ACH processed volume is associated with a 6 to 7.5 percent rise in ACH production costs (last column of table 9). While these results indicate that strong scale economies are associated with ACH processing, these economies are less than those which were implied by looking at the raw data alone (see last row of table 7). One reason for this difference is that the implied measure reflects an assumed linear relationship between average costs and ACH volume while the statistical model specifies a curvilinear relationship. The latter is more correct since a linear relationship can give the counter intuitive result that ACH average costs turn negative at very high volume levels.

As before, SCE estimates significantly different from 1.0, the constant cost case, are starred (*) in table 9. While all the point estimates of the SCE are less than 1.0, not all are significantly less than 1.0. In particular, not all SCEs for ACHs located at Reserve Bank head offices (row 2, table 9) are significantly less than one. This indicates that the standard error for these SCEs is large (it averages .22 over 1980-82). The standard errors computed for the SCEs estimated to exist for all ACH offices together (row 1) and for ACH branch and RCPC offices (row 3) are small—they both averaged .07—and result in SCEs which are significantly less than one for all five years.

For all three sets of SCE estimates, however, the standard errors are always large enough to conclude that the changes in SCEs observed between the different years were *not* significant. That is, any seemingly downward trend in ACH SCEs over time pre and post-pricing is not statistically significant and should not be relied upon. From a statistical standpoint all that can be concluded from table 9 is that significant scale economies exist; there is no significant year-to-year or group-to-group difference in the estimated SCEs. Thus, in contrast to the check results, there appears to have been no change in resource use at ACH offices as a result of pricing. But this may change in the future as more and more costs (and ACH debit float expenses) are reflected in the ACH price. In 1982, only 20 percent of ACH costs

TABLE 9

Scale Economies at Federal Reserve ACH Offices Over 1978-1982

	1978	1979	1980	1981	1982	Five Year Averages
All ACH Offices	.77*	.74*	.80*	.74*	.70*	.75
FRB Head Offices	.70*	.73*	.77	.71	.51*	.68
Branch and RCPC Offices	.60*	.61*	.63*	.59*	.58*	.60

*Indicates that the SCE estimate is significantly different from 1.0 or constant average costs at the 95 percent level of confidence using a two-tailed *t* test (applies to only the first five columns, not to the averages shown for each row in the last column).

were reflected in the price charged. However, this will increase by 20 percentage points each year until 1986 when 100 percent of costs will be reflected in the price and revenues will match all costs (plus the PSAF and ACH float).

Wire Transfer Scale Economies. Estimates of scale economies for wire transfer operations, using cross-section data for 1979, have been presented earlier (Humphrey, 1982) and are only generally discussed here. Since constant average cost was the rule for wire transfers prior to FR pricing. The analysis was not repeated using post-pricing data. In 1979, at only 4 of the 36 wire offices were the scale economy estimates significantly different from 1.0 or constant costs. At that time 98 percent of the wire transfers were processed at offices with constant average costs; only 2 percent were processed at the 4 offices with significant economies to scale, and these had the smallest volume.

When all wire transfer offices were used together, the SCE was 0.97, 1.04, and 1.04 for 1979, 1978, and 1977 respectively. None of these values was significantly different from constant average costs, or 1.0. Applying the statistical model separately to wire operations at Federal Reserve Banks and again to operations at Federal Reserve branches (as was also done in table 9 for ACH operations) yielded numerically lower estimates of SCE for branch operations, and thus higher possibilities of scale economies, but these estimates were not significantly different from those at head offices (FR Banks).

These results seemingly indicate that wire-transfer operations had almost completely "used up" their opportunities to achieve economies of scale prior to pricing and experience constant average production costs. It is very unlikely that in the post FR pricing period that these constant cost conditions have changed. There have been no major technological or organization changes to wire transfer operations which could lead to significant scale economies and if, instead, scale diseconomies were to emerge they would result in a loss of market share until constant cost conditions were again realized (just as occurred for checks when they were priced). The U-shaped average cost relationship implied for wire transfer operations using the raw data in table 7 was also seen prior to pricing but it was not strong enough to generate a statistically significant U-shaped relationship when the scale economies were estimated. Thus it must be disregarded.

Cash and Book-Entry Securities Processing Scale Economies. Two other Reserve Bank payments services have been studied and their scale economies estimated. Coin and currency services, of which only transportation from Reserve Banks to commercial banks is a priced service, were studied by Zimmerman (1981) and a significant

U-shaped average cost was estimated using 1979 cross-section data. Thus scale economies, constant costs, and scale diseconomies appear to exist for the transportation, inspection, and sorting of coin and currency. The inspection for counterfeit bills, destruction of unfit currency, costs of new coin and currency issue, and sorting of monies returned to Reserve Banks were considered by Congress to be services which should be provided to the public at a zero explicit fee and so were not priced. Like checks prior to FR pricing, however, these currently free services seem to be overused from an efficiency standpoint.

Another study by Belton (1984) concerned scale economies in book-entry U.S. government security transfers, of which some 70 percent of the current volume is concentrated at one FR office (New York). Belton found that the book-entry securities transfer operation experienced statistically significant scale economies, whether or not the office with the largest volume was included in the 1982 cross-section data set. The scale economy value estimated was .65 when all offices were restricted to have the same scale economy value. This scale economy result is quite similar to that reported earlier for ACH operations and indicates that considerable cost-efficiencies can be realized in this area if book-entry transfers are expanded in the future.

Scale Economies and Choice of a Pricing Strategy. When scale economies exist, as in ACH and book-entry security operations, the Federal Reserve's choice of a market sensitive pricing strategy can be shown to be optimal. Such a strategy achieves the greatest public benefits because real resources are saved compared to average cost pricing. When constant costs exist, as now occurs for checks and which was evident for wire transfers prior to FR pricing, market sensitive pricing can also be used but the end result should not be any different for society as a whole than if simple average cost pricing were chosen. Using market sensitive pricing when constant costs exist, however, can maintain (or improve) the market shares of FR check and wire transfers when competiting private sector suppliers also use market sensitive (average cost) pricing. Currently, competing suppliers of check services (correspondent banks) also use market sensitive pricing but privately supplied wire transfer services basically use average cost pricing.

Market sensitive pricing, at least the way it was in fact applied to checks, could have misallocated resources when checks experienced scale diseconomies. Correspondingly, the same approach would also be inappropriate if it were applied to Reserve Bank coin and currency services, because of the scale diseconomies experienced there as well. But check processing operations now experience constant costs (after pricing) and, as a result, it is possible to conclude that Federal Reserve services are now being priced in a manner which can be strongly supported in terms of economic efficiency.

C. Retail Payments Operations: Scale Economies in Providing Commercial Bank Deposit and Loan Services

The process of intermediating liabilities, such as deposits and purchased funds, into assets such as loans, security holdings, and other investments is the basic business of depository institutions. The intermediation process is closely tied to the use of various payment services, as when:

1. Check and ACH services are provided to holders of demand deposit and other transaction balances, often in partial payment for bank use of these liabilities;
2. Wire transfer services are used to transfer purchased funds (CDs, federal funds); and when
3. Checks and wire transfers are used to disburse loan proceeds, trade definitive and government securities, and make other investments.

Thus an analysis of scale economies in the provision of commercial bank deposit and loan services also represents an investigation of the cost economies involved in providing retail payment services. These payment services are more broadly defined than those of the Federal Reserve, where only processing, transportation, and settement costs were considered. In this sense, Reserve Banks (and other payment providers) are seen to supply *wholesale* payment services which are "repackaged" and combined with other intermediation operations and thus become, broadly defined, *retail* payment operations—the process underlying funds intermediation in the U.S. banking system. The existence or nonexistence of scale economies in the provision of retail banking services will influence the prices charged for providing payments services both across different sized banks and, if interstate banking is allowed, over time as banks merge with one another and become larger—possibly very large.[44]

Average Costs of Retail Payment and Intermediation Operations. There is no single, generally accepted way to express the average costs of bank deposit and loan activities. This is because there remains disagreement about how bank output is most properly measured. Two approaches to viewing bank output or the financial intermediation process can be distinguished.

---

[44] The analysis in this section is discussed in more detail in Humphrey (1984) and Benston, Hanweck, and Humphrey (1982).

1. (1) The Production Approach: this approach relates operating costs to the *number of accounts* serviced through processing deposit and loan documents as well as debiting and crediting these accounts; and
2. (2) The Intermediation Approach: this approach focuses on the operating and interest costs incurred to intermediate the *dollars* in deposit accounts into loans.

The production approach is analogous to methods used to investigate scale economies in Federal Reserve payments processing operations or, more generally, the scale economies of manufacturing firms where emphasis is on the production and measurement of output in physical units like the weight of the output or numbers of units produced. Thus the output produced by a financial institution would here be viewed as the number of accounts produced and/or serviced through expenditures on operating costs.

The intermediation approach argues that, while banks do indeed produce and service deposit and loan accounts, this production process is today more closely associated with the operating and interest costs incurred per dollar in these accounts. Interest costs are here viewed as a direct payment for deposit funds. This direct payment should be considered along with the indirect or operating cost payments to depositors reflected in the type of services supplied in exchange for the (idle or low interest earning) balances provided. In addition, purchased funds should be included in the dollar measure of output. Purchased money has been increasingly used to fund assets and differs only from other deposits through the ratio of direct (interest) to indirect (operating cost) payments provided to balance holders. In sum, the two different approaches to measuring bank output imply the following definitions of output and cost for a study of retail bank scale economies:

	Production Approach	Intermediation Approach
Bank Output	Numbers of deposit and loan accounts.	Dollars in deposit and loan accounts, including purchased funds.
Total Costs	Total operating costs.	Total operating and interest costs.

Although different, either way of viewing bank output and its associated costs is basically correct but the scale economy results need not be identical. For example, although scale *disecono-*

mies may exist at larger banks in the direct production of the number of (say) deposit accounts, the extra cost per account is acceptable if it is fully (or more than fully) offset by having proportionally more dollars in the account. In this sense, diseconomies of scale per account produced using the production approach could coexist with scale economies per dollar intermediated using the intermediation approach. Fortunately, it turns out that such direct conflicts in scale economy results are not often observed. Although the numerical values differ, similar scale economy conclusions are obtained using either approach.

This result is shown implicitly in table 10 where the average unit costs of producing five bank services—demand deposits, time and savings deposits, installment loans, real estate loans, and commercial and industrial loans—have been computed from cost accounting data for 625 banks in 1981. These average costs for the aggregate of 5 banking services are shown for the production and intermediation approaches and separately for banks in unit and branch banking states. The clear conclusion from table 10 is that, regardless of the bank output definition used and regardless of organizational form (unit or branching), the average cost curve for retail payments operations is U-shaped, indicating that banks first experience falling, then constant, and finally increasing costs as output is successively increased.[45]

The U-shaped nature of the average cost curve is perhaps more easily seen when the unit and branch bank average costs are combined for the production and intermediation approaches and expressed graphically, as in figure 3. While both approaches produce a U-shaped average cost curve, the intermediation approach (dashed line, right scale) produces a much flatter curve, indicating weaker scale economies and diseconomies than that which would be associated with the production approach (solid line, left scale).

Scale Economies in Commercial Banking. Exact measures of bank retail payment and intermediation operation scale economies are not presented here since these statistical results closely mirror the results seen in table 10 and figure 3. In brief, while scale economies exist in retail payment and bank intermediation operations, they are limited to smaller commercial banks, typically those not exceeding $50 to $75 million in deposits. This holds using either the production or intermediation approach to measuring bank output. The very largest banks, in contrast, either experience very strong scale diseconomies (using the production approach) or have constant costs to slight scale diseconomies (using the intermediation approach).

---

[45] The U-shaped curve exists even when average costs are alternatively expressed as (total operating costs)/(dollars in the accounts) or as (total operating and interest cost)/(number of accounts). These computations are not shown here since they inappropriately mix the two appraches to measuring bank output; they were developed only for completeness and to show the robustness of the U-shape cost curve result.

The important implication for commercial bank pricing of payment services is that larger banks, despite their presumed easier access to newer technology and possible superior organizational structure, do not on average experience lower operating costs or operating and interest costs per unit of output produced than do smaller banks. As a result, changes in banking structure through mergers or acqusitions (e.g., interstate banking) should not lead to lower payment costs to final users because scale economies do not appear to exist in commercial bank retail operations as they are now configured.

These particular results are different from the conventional wisdom in banking, reinforced by earlier bank scale economy studies, that scale economies exist for all size classes of banks. There are at least two reasons why our newer results reject scale economies for larger banks:

TABLE 10

Average Cost (per account and per dollar) of Five Banking Services (1981)

	The Production Approach: Average Total Operating Costs Per Account (dollars per account)		The Intermediation Approach: Average Total Operating Plus Interest Costs Per Dollar In The Accounts (cents per dollar)	
Deposit Size Class: ($ millions)	Unit State Banks	Branch State Banks	Unit State Banks	Branch State Banks
<10	$130	$65	9.0¢	7.8¢
10-25	70	*59*	8.2	*7.7*
25-50	*38*	62	*8.1*	8.0
50-75	68	68	8.5	7.8
75-100	78	66	8.5	8.2
100-200	96	64	9.7	8.4
200-300	105	71	10.0	8.8
300-400	103	75	9.7	9.3
>400	192	99	11.3	9.5
Average	71	76	9.2	8.8

Source: Computed from bank cost accounting data covering a sample of 625 banks in Board of Governors of the Federal Reserve System, *Functional Cost Analysis*, 1981.

Note: Values in italics show where average costs reach their minimum point. The 5 banking services are demand deposits, time and savings deposits, real estate loans, commercial and industrial loans, and installment loans. These 5 banking services account for over 70 percent of total operating costs.

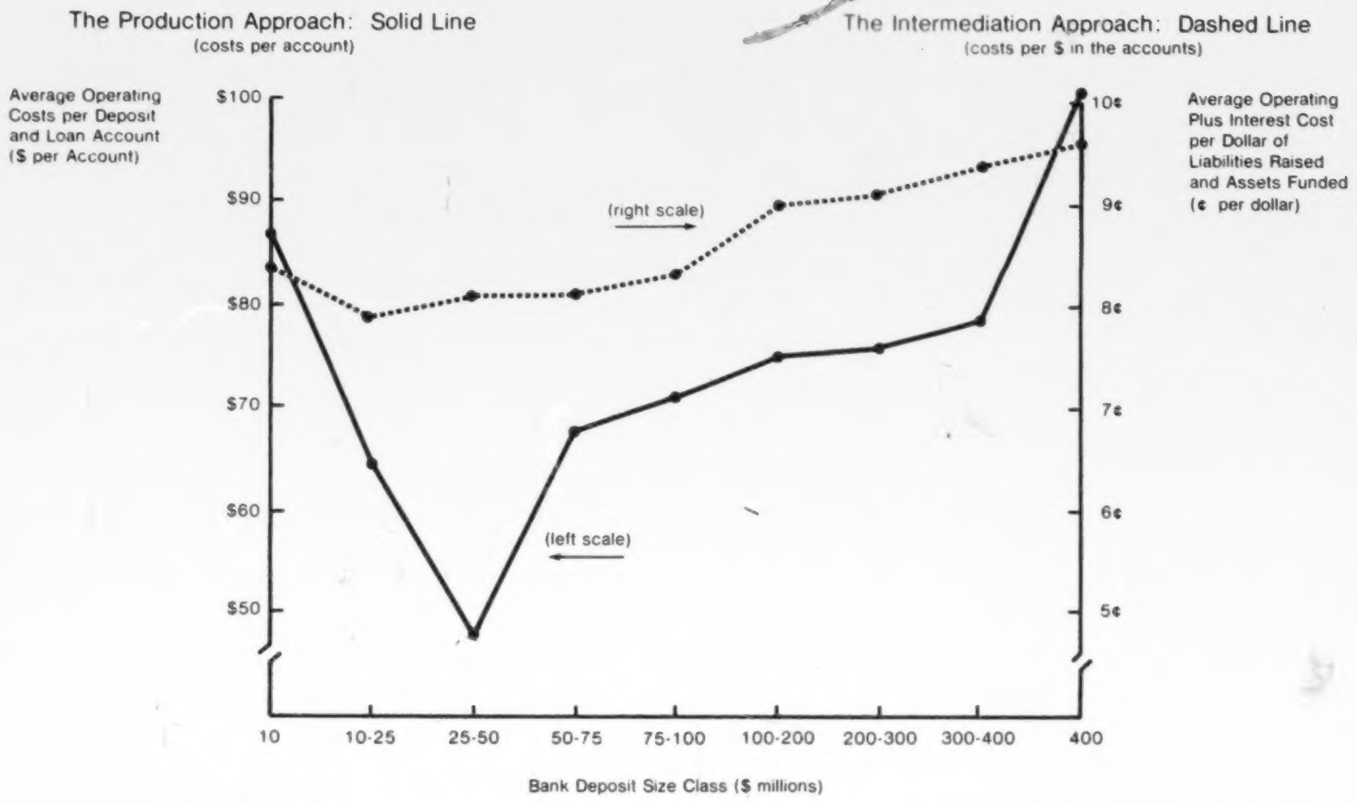
(1) The earlier bank scale economy studies typically did not use the most appropriate statistical model in their analyses, one which could in fact identify scale economies at small banks and diseconomies at larger banks if they existed simultaneously in the data; and

(2) The measurement of scale economies for branch banking organiations, even if the right statistical model is used, was not done correctly and this unfortunately biased the empirical results toward finding scale economies when they did not exist for the bank as a whole.

These assertions are more fully discussed and defended elsewhere (Humphrey, 1985) but the common sense of our result is clearly seen in the U-shaped relationship pictured in table 10 and figure 3 using the raw data alone, results which are supported in the statistical analysis which derived the explicit bank scale economy estimates.

FIGURE 3

Average Cost Curves by Bank Size Class: The Production and Intermediation Approaches to Measuring Bank Output
(1981: Unit and branch state banks together)



VII. Competition in Providing Wholesale Payments Services

Competition between banks in deposit markets is strong and relatively unrestricted when deposits (CDs, federal funds) are purchased directly from sellers. The real restriction on interbank competition is in physically accepting deposits within unit banking states or across state lines. This restriction also applies to other banks or depository institutions which deposit checks (cash letters) at correspondent banks for processing, collection, and settlement and also seemingly applies to the deposit of ACH tapes (but not to the telephonic or computer initiation of wire transfers). Federal Reserve Banks are not so restricted since their 48 offices are located in almost every state. In effect, Reserve Banks comprise the nation's only current interstate banking organization which can provide nationwide *wholesale* payment services directly to local users.

A. Existing Competitive Structure

Check Market Share. Substantial private competition already exists in check processing. Of the estimated 40 billion checks written in 1983, some 12 billion (30 percent) were deposited at the same bank they were drawn upon. These "on us" items therefore required no further processing. Of the remaining 28 billion checks, 15 billion (37 percent) were processed by correspondent banks and local clearinghouses while 13 billion (33 percent) were deposited at the Federal Reserve. The involvement of the private sector in check processing was thus substantial: over 67 percent of all checks written were processed entirely by commercial banks, other depository institutions, and private clearinghouses. The Federal Reserve's market share for wholesale check processing services is estimated to be 33 percent, as shown in table 11.

In 1979, prior to Federal Reserve pricing, the Federal Reserve's market share was estimated to be around 43 percent. Thus Federal Reserve pricing has, as one might expect when prices rise from zero to some positive amount, altered relative prices in the favor of private sector suppliers and led to a roughly 25 percent reduction in the FR market share for processed checks. More importantly, pricing has meant that users of check processing services now face prices which reflect the resource costs involved in providing this service. As was

TABLE 11

Volume and FR Market Share in Check, ACH, and Wire Transfer Wholesale Payments Services (1983)

	FR Volume (1983; millions)	Total Volume (1983; millions)	FR Market Share (1979)	FR Market Share (1983)
Check	13,000	40,000	43%	33%
ACH	380	400	95%	95%
Wire Transfer	38	57	64%	67%

Sources: Data derived from the same sources used in tables 1 and 2.

seen above, resource allocation has been improved as a result of FR pricing.

Changes in check market share in the near future will likely be determined through the development of new check payment "products", such as check truncation, electronic check collection, and a proposed direct settlement service. These are discussed below. The various suppliers, at present, use the same check "production functions" (e.g., the same reader-sorters, similar transportation modes, etc.) so that neither one of them appears to have any inherent cost advantage over the other now that FR pricing is in place. The same conclusion may not apply to ACH, however.

ACH Market Share. Except for New York, the Federal Reserve processes virtually all ACH items; its market share is around 95 percent both before and after FR pricing. But while scale economies exist in ACH processing, seemingly indicating that new suppliers of ACH services would possibly be at a competitive disadvantage, this does not seem to be the case. New private sector suppliers are in fact entering the ACH processing market and the Federal Reserve's market share is expected to fall in the near future.

The main reasons why new entry is occurring in the ACH area, even though scale economies were demonstrated to exist above, concerns Federal Reserve pricing practices and the ACH production function. Specifically:

(1) Reserve Bank market sensitive prices for particular ACH services within the ACH service line did not (prior to 1984) reflect the fact that unit costs fall as user volume rises; and

(2) The average cost of providing the ACH service, even though falling as volume increases, may not be as low as it could be due to the current unconsolidated structure of FR ACH production.

The importance of incorporating "volume discounts" into the ACH market sensitive pricing structure is greatly increased when different users demand markedly different amounts of output, as occurs in the ACH area. For example, commercial ACH volume is generated by only around 1,000 originators but there are over 20,000 receivers. Originators (of debits or credits) are typically large banks while receivers (of debits or credits) are smaller bank and thrift institutions. While every originator is also a receiver, 95 percent of the receivers do not originate. Overall, the most active and largest 5 percent of ACH market participants account for nearly one-half of total commercial ACH volume. All of government ACH transfers (about 60 percent of total ACH volume in 1984) represent volume originated by a single originator going to a very broad receiver base (e.g., government and military salary and retirement payments, social security payments, etc.). Thus ACH volume is concentrated at a relatively small number of high volume originators (who are also, for the commercial volume component, the highest volume receivers). These high volume users are precisely those who should reap the benefits of scale economies in Federal Reserve ACH processing. This is accomplished by creating market sensitive prices for a larger set of specific user groups. If such user targeting is not done, Reserve Banks can expect to lose volume to private sector suppliers whose prices more accurately reflect the scale economies involved. This is an argument for more correctly pricing the ACH service, so that the *relative* prices to different users better reflect the actual relative costs involved. But having the proper relative user price is not the whole answer; one needs to compare the *absolute* FR price with the price charged by alternative private sector suppliers. The absolute FR ACH price is basically determined by the level of input costs, the type of ACH output produced (e.g., electronic or magnetic tape), and the degree of ACH office consolidation.

Reserve Banks currently operate 35 separate ACH offices rather than one or two very large processing centers. By operating so many separate processing centers, Reserve Bank ACH scale economies are not as fully realized as they would otherwise be if all operations were much more consolidated.[46] Thus the average unit cost of an unconsolidated ACH processing system (operated by Reserve Banks) is likely to exceed that of a fully consolidated one (operated by new private sector entrants).

The Federal Reserve's ACH production structure was set up prior to pricing. Each FR District, and often each FR branch office, had ACH processing capability, mirroring the check production structure. Investments were made in the current unconsolidated processing structure which still have useful lives. As a result, consolidation could very well be uneconomic in the short-run but cost-effective in the long-run when this capital is fully depreciated. The new ACH entrants do not have either of these limitations since they already are consolidated and are not organized in the same decentralized manner as is the Federal Reserve System.

---

[46] In a consolidated ACH environment, extra data communications costs between FR offices would be incurred. This would offset some of the pure scale economy benefits of office consolidation.

Wire Transfer Market Share. Like checks, direct competition in wire transfers is significant. The four private sector wire networks which, along with FedWire, process U.S. bank wire transfers had a combined volume of 57 million in 1983 worth more than $142 trillion over the year.[47] FedWire volume was 38 million transfers (removing double counting) worth $85 trillion. In terms of processing volume, the Federal Reserve's market share in 1983 was thus 67 percent.[48] In 1979, prior to pricing, FedWire's market share was estimated to be around 64 percent, so the change in share as a result of pricing has not been significant.

New entry into the wire transfer service area has been limited after pricing in part because of the restrictions placed on access to the net settlement service in order to reduce payments system risk. Once the net settlement access issue is finally and fully resolved, it is expected that new local/regional large dollar wire transfer networks will emerge in other money centers in addition to those private networks now in operation in New York and Chicago.

Perhaps a longer-run threat to FedWire volume lies in the possibility of shifting smaller dollar wire transfers to ACH processing once same-day ACH settlement is available (likely sometime in 1985). There are some 8,500 institutions which use FedWire and one-half of them are on-line with Reserve Banks (the other half use the telephone to initiate a wire transfer).[49] Some high volume institutions have on-line capability with both wire transfers and ACH and this dual capability is likely to increase in the future. Thus with same-day ACH settlement, as exists for wire transfers, it is possible that small dollar wire transfers—which cost around $.65 to initiate—will shift to ACH transfers—which now cost less than $.05 to initiate. Small dollar wire transfers, those for less than $100,000, account for around 58 percent of total FedWire volume so the potential for the ACH to absorb FedWire volume is quite real. In addition, commercial bank wire transfer ($5.20) and ACH prices (4.6¢), as shown in table 3, also favor such a substitution for end users.

---

[47] The wire transfer networks used here, in addition to FedWire, were CHIPS, BankWire (CashWire), and CHESS. CHIPS is far and away the largest private wire network in the U.S.

[48] Reserve Banks do not offer an international wire transfer service whereas around 90 percent of CHIPS volume is international. Another international network is SWIFT, with about 104 million transfers worldwide in 1983. The Federal Reserve's market share of worldwide wire transfers (i.e., including SWIFT's transfers) is 24 percent. The market share figure when 90 percent of CHIPS and all of SWIFT's transfers are excluded is 95 percent. The market share figures in the text included all of CHIPS transfers and thus refer to wire transfers where one or two U.S. banks are parties to the transfer.

[49] Each of the 12 Federal Reserve District head offices has wire transfer capability—both on-line and telephonic. The telephone component, however is the most expensive per transfer. This component could likely be cost-effectively consolidated at one national U.S. central office using an 800 phone number. Even if all institutions were on-line, there would still be a need for telephonic initiation of a wire transfer, as when computer problems arise. Almost every operating day, different on-line users experience computer downtime problems in their wire transfer area and use the telephone as a backup device.

B. The Importance of Quality and Other Product Characteristics

The quality of a payment service affects user demand along with price. The speed of funds availability is the most important aspect of check processing quality. Priority processing of checks drawn on high volume presentment endpoints was initiated by Reserve Banks in 1983 to:

- Provide depositing banks with better funds availability;
- Reduce Federal Reserve float costs; and
- Limit paying bank and corporate cash managers' incentives to engage in remote disbursement.

At this writing, it is too early to determine the success of this program—known as noon presentment and high dollar group sort in the industry—in improving the quality of the FR check service.

Quality in the ACH service has also concerned funds availability since payments information would often have to be received at a Reserve Bank one to two days prior to when the transfer was to occur. A night processing cycle reduced the stringency of this requirement and, in effect, opened up ACH usage to cash management functions previously handled by checks. Another ACH quality consideration is message format. Allowing payment invoice information and other internal corporate accounting data to be included along with traditional ACH payment information is currently being tried in the ACH Corporate Trade Payments program.

Message format is also an issue in the wire transfer area. This is because most networks have different formats. While a standardized third party message format has been recommended by the ABA and supported by the Federal Reserve, it has not yet been required by all networks (including FedWire). The major stumbling block is that only the automated users of wire transfers—basically the largest banks with on-line accounting systems—stand to benefit from message format standardization. In contrast, all users will have to incur the costs of modifying their internal processing systems to send and receive the standardized format. Thus nonautomated banks incur costs without gaining an equivalent benefit while the role is reversed for automated banks.

C. Potential New Payment Services

A number of potential new payment services are or have been under investigation or are now in the pilot stage. These new services basically augment the existing services and concern:

(1) Checks—check truncation, electronic check collection (ECC), and a direct settlement service (DSS);
(2) ACH—Value dating and same-day settlement; and
(3) Wire transfers—value dating.

Check truncation. Check truncation occurs when the physical check is kept at the bank of first deposit or at some other point in the collection process so the typical movement of the physical item is truncated; certain processing costs may be saved as a result. Credit unions are the largest users of check truncation and in this application the physical item is kept by the paying credit union or its paying correspondent. If all transaction accounts were handled in this manner, then the truncation of the physical item could of course be shifted backward and occur at the bank of first deposit with payment information transferred to the payor institution by electronic means.

Electronic Check Collection (ECC). A closely related concept is electronic check collection where either the payment information on a check or an electronic image of both sides of the check itself are transmitted to paying banks for collection. But while check truncation would apply to all checks written, a Federal Reserve ECC proposal envisioned that only large dollar checks would be handled electronically. In addition, there would be no truncation as the physical item would still be presented using existing transportation networks.

ECC offers two benefits regardless of who supplies the service:

(1) ECC would speed the collection of large dollar checks and reduce incentives for payor banks and corporate cash managers to spend real resources on remote disbursement (thereby removing a clear social inefficiency in the payments mechanism); and,
(2) ECC would allow the supplier of this service to substantially improve funds availability to depositing banks; and, by substituting electronic delivery for expensive expedited physical delivery now reserved for large dollar checks by correspondents, save real resources here as well.

Offsetting these potential benefits would be the extra costs of electronic capture of the payments information and electronic receipt of this information at payor banks. There may also be an increased risk in paying forged large dollar checks.

In the Federal Reserve's analysis of these costs and benefits, the float benefit to depositing banks was offset by the float loss at paying banks. Thus the analysis undertaken concerned *social* costs and benefits. This analysis appropriately excluded the float benefits which would accrue to Reserve Banks (and be passed on to banks depositing large dollar checks with the Federal Reserve rather than with correspondents). The Federal Reserve's analysis suggested that the social costs outweighed the social benefits and for this reason, along with certain legal considerations, the ECC proposal was not adopted.

Direct Settlement Service (DSS). Correspondent banks have requested that the Federal Reserve allow them to act, in effect, as agents of Reserve Banks when collecting checks. Under DSS a corresondent bank would process and transport checks directly to a paying bank. But, in contrast to what commonly occurs today when this is done without bilateral agreements between collecting and paying banks, the presenting correspondent bank: (a) would not face any presentment fee for presentments outside of local clearinghouse hours; and (b), would be paid in immediately available funds by having Reserve Banks credit the collecting bank's reserve account and debit that of the paying bank. It is argued that DSS, if offered by the Federal Reserve, would serve to remove the advantages which accrue to Reserve Banks through law and regulation. That is, Reserve Banks by law can not be charged a presentment fee and, by regulation, must be paid in immediately available funds (rather than by check, for example).

Presentment fees are typically levied on correspondent banks which present items to be paid outside of clearinghouse hours. This fee, usually 1.5 to 4 cents per check, was meant to cover the extra accounting costs incurred by paying banks when items are paid. Its effect, of course, is to discourage direct presentment by other banks and limit presentments to those early to late morning hours deemed by local banks to be the time they will accept checks drawn on one another. The accounting costs involved when items are presented at clearinghouse hours are waived because they would tend to offset each other for the clearinghouse members.

Charging a presentment fee for presentments at other than clearinghouse hours enables paying banks to obtain more float than if presentments were to always occur later in the day. This is because more checks could be processed for presentment late in the day for all days. Legally, presentment can occur up to 2:00 p.m. in the afternoon;

however, the law is silent on presentment fees for other than Reserve Banks.

Federal Reserve analysis of DSS suggested that if it were offered, correspondent banks would have inducements to spend real resources up to the point where their extra processing and transportation expenses involved would equal the extra revenues derived from reducing float. This is a result of equating private costs and benefits where any float reduction is viewed, quite correctly here, as a gain to the collecting bank. In the social cost/benefit analysis of the Federal Reserve, float is merely a redistribution from paying to collecting banks and is properly excluded.

But there would be a social benefit to DSS—a benefit not included in a collecting bank's cost/benefit calculation. The social benefit would come from reduced expenditures of real resources for remote disbursement since the float gain from remote disbursement would be reduced under DSS. The Federal Reserve study suggested that, if correspondent banks pursued DSS according to a private cost/benefit calculation, the corresponding social cost/benefit ratio (which neglects float but includes the reduction in remote disbursement) would show more real resources being spent than are saved. While it can be argued that applying a social cost/benefit "test" to a private correspondent bank is inappropriate, these same "tests" have been applied in the past to Federal Reserve payments initiatives. ECC failed such a test while the recently adopted high dollar group sort program passed it.

Like ECC, DSS could have important competitive effects on correspondent bank/Reserve Bank market shares in check processing operations. Simply put, industry observers have suggested that ECC would have favored Reserve Banks while DSS would favor correspondent banks. The DSS issue has not yet been resolved.

Value Dating and Same-Day Settlement. Value dating in an ACH or wire transfer environment would permit originators the flexibility to submit funds transfer messages days before the message would actually be sent. Originators could specify the date upon which the message has value, rather than be constrained to follow rigid deposit deadlines for ACH payments or same-day requirements for wire transfers. Value dating capability is built into the ACH but has not yet been adopted for wire transfers although the idea itself is not new.

A related concept is same-day settlement where funds are transferred and settled the same day that the payments message is deposited at an ACH. Wire transfers already have same-day settlement. As noted above, once the ACH has this capability, it is likely that some small-dollar recurring payments now processed as wire transfers will shift to the ACH.

VIII. Risk in Payments Systems

Risk exists in all payments systems but, by common consent, it appears to be greatest on large-dollar wire transfer systems. These systems together transfer up to $500 billion a day and have an average transaction value of $2.5 million per transfer. For checks, ACH, and cash payments, the average amount per transfer is less than $1,000 and the total amount transacted per day less than $150 billion. Thus checks and ACH transfers are basically small-dollar payments systems and the discussion in this section focuses on the large dollar payments made over wire transfer networks. Specifically, the issue of *daylight overdrafts* and the potential credit and systemic risks these overdrafts pose to the operation of large-dollar networks are discussed. If a participant with a daylight overdraft is unable to settle on a large-dollar network, Reserve Banks (on FedWire) and net receivers of funds (on private networks) can suffer a loss. While the probability of an unexpected failure to settle is low, the amount of such a loss could be large since the daylight exposures involved are themselves large.

As an illustration of the systemic risks involved on one payments network (CHIPS), the inability to settle of a major participant was found likely to lead to the failure to settle of almost one-half of the over 100 bank participants on the network and resulted in over one-third of that day's payments being cancelled. Similar results were obtained when the simulation was repeated on a different day and for simulations where a different participant was assumed to be unable to settle. These simulations, while only illustrative and performed prior to current CHIPS risk reduction efforts using bilateral net credit limits, suggest that systemic risk on large dollar payments networks can be significant and that its possible effects on financial markets may be large enough to be a serious problem if a failure to settle were to in fact occur.

A. Daylight Overdrafts on Large-Dollar Funds Transfer Networks

Fedwire Daylight Overdrafts. Daylight overdrafts occur on the Federal Reserve's FedWire wire transfer network when funds sent so exceed funds received that the sender's reserve account has a negative balance during the day. An illustrative daylight overdraft on

FedWire is shown in figure 4. Based on Federal Reserve surveys taken in 1979 and later in 1981, it was demonstrated that the reserve account balances of banks incurring daylight overdrafts typically follow a definite pattern during the day. Banks almost always open the business day with positive reserve balances but draw them down into an overdraft position by about 11 a.m. Eastern time. The overdraft tends to remain, fluctuating in amount, until just before 3 p.m. Eastern time when it is erased by incoming flows. This cycle is reasonably consistent across large banks and indicates that the problem is more than just a minor and short-lived imbalance in funds flows or a result of computer downtime problems at banks or in the Federal Reserve. The 11 a.m.-3 p.m. period covers 80 percent of all maximum overdraft occurrences (a 10 a.m.-4 p.m. range covers 95 percent of the maximum occurrences).

Between 1979 and 1981, FedWire daylight overdrafts rose by almost 50 percent at large banks; FedWire volume rose by less than 45 percent. Large banks, those with deposits in excess of $1 billion or 2.6 percent of all institutions using FedWire, accounted for 91 percent of the dollar value of all daylight overdrafts. Although the average value of daylight overdrafts for all overdrafting institutions was $61 million, it averaged over $325 million per day for overdrafting banks with deposits in excess of $1 billion.

One "cause" of FedWire daylight overdrafts appears to be the current institutional arrangements for handling overnight federal funds transactions. Total federal funds transactions are estimated to have accounted for 44 percent of FedWire dollar volume in 1981. Daylight overdrafts, in contrast, were around 12 percent of FedWire dollar volume during the same period. Many federal funds purchases, although negotiated in the morning, result in an inflow of funds to the borrowing bank late in the same day. These borrowings are usually repaid early in the next business day (less than 24 hours after the funds were received). As a result, a large morning outflow of funds occurs from banks relying for funding on overnight liabilities and a similar reverse flow takes place in the afternoon. The cycle has evolved because, in the absence of a policy that forbids or penalizes daylight overdrafts, there is currently little incentive to maintain a positive reserve account balance throughout the business day.

CHIPS Daylight Overdrafts. Daylight overdrafts on CHIPS are often as large as those which occur on FedWire. These overdrafts differ from FedWire overdrafts in three small, but important, respects:

(1) There are no idle balances held by CHIPS participants to reduce the value of a CHIPS overdraft. Idle reserve balances,

FIGURE 4

Illustrative Reserve Account Balance by Time of Day for an Institution with a FedWire Daylight Overdraft

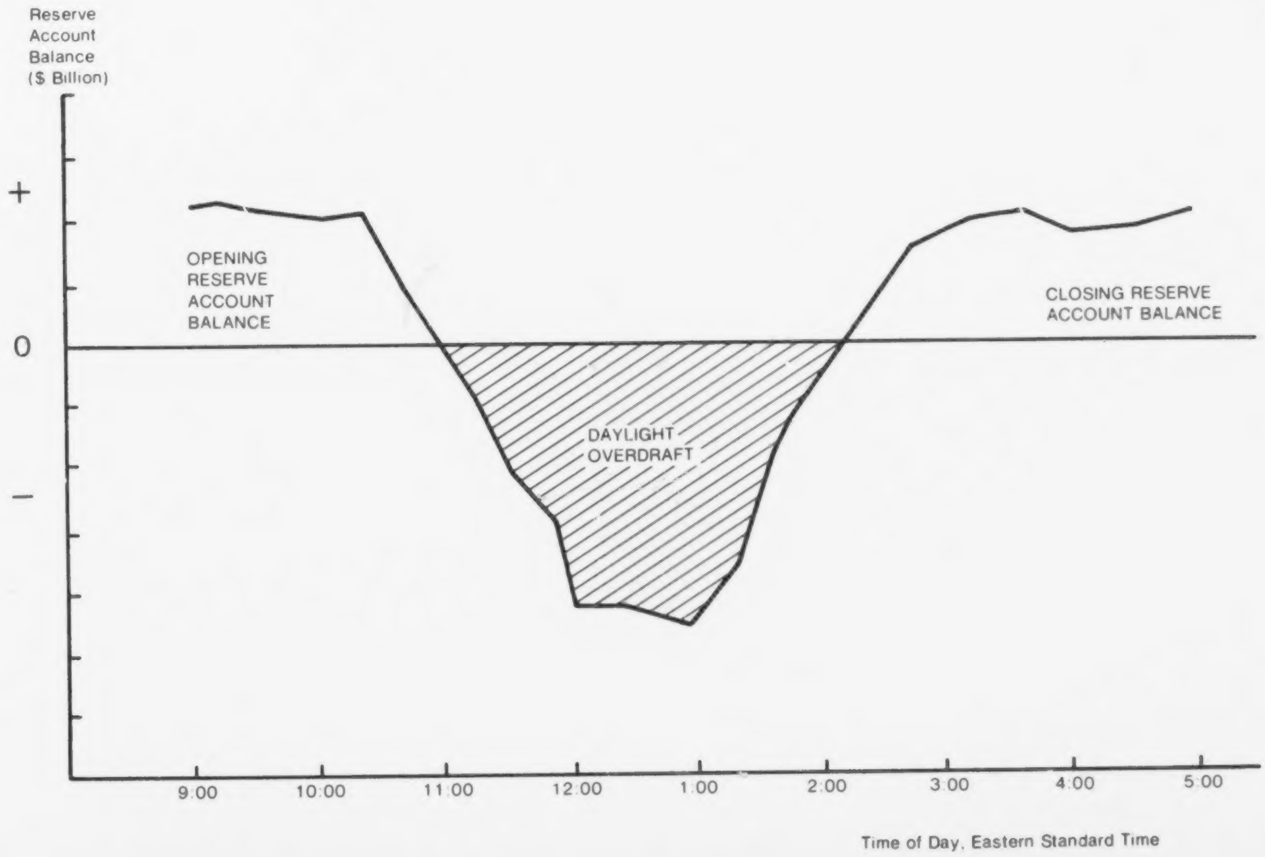


TABLE 12

Actual Funds Transfer Daylight Overdrafts on FedWire and CHIPS
(January 1983, $ millions, 21 days)

Bank	Average Daily Overdraft	Number of Overdrafts	One Day Extreme Overdraft
1. FedWire	$ 546	9	$1,520
CHIPS	1,283	21	2,474
2. FedWire	4,169	21	6,096
CHIPS	1,033	21	1,669
3. FedWire	1,327	17	2,649
CHIPS	2,745	21	4,519

Source: Internal Federal Reserve daylight overdraft survey data, January 1983. Data on individual banks by name of bank can not be released. Thus individual banks are not herein identified. Note that U.S. government book-entry security transfer daylight overdrafts are excluded from these figures.

however, reduce the value of a FedWire overdraft compared to what it would otherwise be;

(2) The time of peak CHIPS daylight overdraft occurs, on average, about one hour sooner in the morning than the peak FedWire daylight overdraft; and,

(3) While FedWire daylight overdrafts generally follow a cycle (see figure 4), CHIPS overdrafts often occur as automated payments instructions received during the (eight hour earlier) business day in Europe are sent in the morning in the U.S. The sending institution often remains in overdraft on CHIPS until end-of-day net settlement occurs, rather than having the overdraft erased by purchased funds inflows as on FedWire.

Unlike FedWire, where the cause of daylight overdrafts appears to be closely associated with the current institutional arrangements in the overnight federal funds market, the proximate cause of CHIPS daylight overdrafts has not been subjected to close analysis and indentified. In the event that CHIPS overdrafts are found to be related to overnight Eurodollar borrowings and/or to the payment of foreign exchange contracts, it may be possible to ''rearrange'' the institutional structure of this market to reduce overdrafts, as is discussed below for federal funds transfers on FedWire.

Daylight overdrafts are often large and have been regular events for at least some large bank CHIPS and FedWire participants. This is seen in table 12 where the funds transfer daylight overdraft activity of three large banks are shown for 21 days in January 1983. The average daily daylight overdraft, for those days when an overdraft occurred, ranges between $546 million to $4.2 billion. The daylight overdraft value used here and elsewhere always refers to the maximum overdraft on a single day (i.e., the lowest point on the line shown in figure 4). Table 12 shows the largest one day funds transfer overdraft experienced during the survey period for the selected banks. This one day maximum figure often is twice the average amount shown. As seen, some banks incurred daylight overdrafts on CHIPS or FedWire on each business day during the 21 day survey period.

Other Large Dollar Networks. There are at least four additional wire transfer (or message transfer) networks which can experience daylight overdrafts. These networks are BankWire, SWIFT, CashWire, and CHESS. Both BankWire and SWIFT are message transfer networks which transfer funds between accounts internally at a single institution. If the balances in these correspondent accounts are insufficient, an *internal* daylight overdraft can occur and the balance is replenished by other internal transfers over BankWire or SWIFT or through *external* wire transfers using FedWire or CHIPS. The CashWire and CHESS networks are relatively small and only end-of-day net settlement information on net debit positions are available, rather than daylight debit or overdraft information as exists for CHIPS and FedWire. Consequently, it is not possible to indicate the average daylight overdrafts which are incurred on these four smaller wire transfer networks but their amounts are believed to be much, much lower.

Reducing FedWire Daylight Overdrafts through Reductions in Overnight Funding.[50] It is possible to estimate directly the extent that FedWire funds and securities transfer daylight overdrafts could be reduced if overdrafting banks shifted more of their overnight federal funds transactions to a multi-day or continuing contract basis. This estimate is obtained from computing the ratio of funds and securities daylight overdrafts (OD) to net federal funds purchased (NFF) at individual large banks which run overdrafts:

OD/NFF = (daylight overdraft on one day)/(net federal funds purchased on the prior day)[51]

This ratio shows the percentage that the daylight overdraft observed on each of the 20 days in a 1981 survey period is of net federal funds purchased on the day preceding the daylight overdraft. Computations for the 114 banks with deposits in excess of $1 billion and incurring daylight overdrafts yielded the results shown in table 13.

---

[50] This discussion closely follows material released by the Federal Reserve to the public through a freedom of information request. As the original author of this material, I took the liberty of shortening the discussion without altering the essential issues involved.

[51] Net federal funds purchased is defined as federal funds and RPs purchased less federal funds and RPs sold. This is positive for large banks who are large net purchasers. NFF is reduced by 25 percent to account for the estimated current existence of multi-day and continuing contract federal funds transactions which do not contribute to each day's daylight overdraft amount.

Sixty-one of the 114 large banks had daylight overdrafts which were 25 percent or less of their prior day's federal funds net purchases. These banks contributed 11 percent of the total dollar value of all daylight overdrafts experienced at large banks in 1981 and comprised 33 percent of the (gross) federal funds market.[52] If only one-fourth of the federal funds transactions at these banks were shifted to a multi-day or continuing contract basis, their daylight overdrafts would disappear, and the total value of large bank daylight overdrafts experienced would fall by 11 percent. If 10 additional banks with daylight overdraft/net federal funds purchased ratios between 25 and 33 percent shifted one-third of their federal funds transactions to a multi-day or continuing contract basis then the total daylight overdraft value at all large banks would fall by an additional 9 percentage points. Overall, if 61, 10, 15, and 28 banks shifted (respectively) one-fourth, one-third, one-half, and all of their current federal funds transactions into multi-day or continuing contracts (or even to internal borrowings on a book-entry basis), 81 percent of the dollar value of daylight overdrafts at large banks would be eliminated.

These estimates also indicate the extent to which existing federal funds transaction relationships would be "rearranged" in order to eliminate daylight overdrafts at large banks. For example, shifting one-fourth of existing (daily) federal funds transactions at 61 banks to a multi-day or continuing contract basis will affect 8 percent of the large bank component of the federal funds market (8 percent equals one-fourth of 33 percent, the percent that federal funds transactions at these 61 banks are of federal funds transactions).[53] Overall, a movement to multi-day or continuing contracts by 114 banks would affect 33 percent of all large bank federal funds transactions.[54] If daylight overdrafts were limited or if charges were imposed on these overdrafts, it is possible that overdrafting banks would shift to increased use of continuing or multi-day contracts and, at a maximum, 33 percent of existing (daily) large bank federal funds volume would be "rearranged."

---

[52] The federal funds market referred to here is restricted to the 122 banks for which daily federal funds data are available. These banks accounted for around 72 percent of all federal funds transacted at all institutions (as shown in the *Consolidated Report of Condition* or *Call Report*).

[53] The "rearrangement" of the federal funds market is somewhat overstated here. The computation assumes that all institutions within the 25 percent or less category in table 13 would in fact have to shift one-fourth of their federal funds. In fact, institutions with lower ratios (e.g., 15 percent or 20 percent) would only have to shift what their ratios were, not the upper bound to the range shown in the table (25 percent).

[54] The 33 percent figure was obtained from .25 x (33 percent) + .33 x (7 percent) + .50 (13 percent) + 1.00 x (16 percent) or the multiplication of the elements in the last column in table 13 by corresponding elements in the first column.

TABLE 13

Daylight Overdrafts as a Percent of Federal Funds Transactions and Reduction in Daylight Overdrafts if Multi-Day or Continuing Contracts were Used

Daylight Overdrafts as a Percent of (Non-Continuing Contract) Federal Funds Purchased—OD/NFF	Reduction in Daylight Overdraft $ Value if Used Multi-Day or Continuing Contracts	Number of Affected Banks with $1 Billion or More in Deposits	Federal Funds at Affected Banks as a Percent of Federal Funds Transactions Reported on FR-2415
25% or less	11%	61	33%
25%-33%	9%	10	7%
33%-50%	17%	15	13%
50%-100%	25%	14	10%
>100%	19%	14	6%
Cumulative Total:	81%	114	69%

NOTE: 91 percent of the total dollar value of daylight overdrafts at all banks were accounted for by banks with deposits of $1 billion or more. The net federal funds (NFF) values used to derive the percentages in column 1 have been reduced by 25%. This reflects the existence of multi-day and continuing contract federal funds transactions which do not contribute to each day's daylight overdraft amount in the numerator.

Since large bank use of federal funds represent around 72 percent to total federal funds dollar volume at all institutions, the shift to continuing or multi-day contracts would only affect 24 percent of the entire federal funds market (24 = 33 × .72). Presently, 25 percent of the total federal funds dollar value is estimated to be transacted under continuing or multi-day contracts. Doubling this percentage to 50 percent would thus be required to eliminate 81 percent of the dollar value of funds and securities daylight overdrafts at large banks.

This calculation does not reflect the reduction in daylight overdrafts which would occur if book-entry U.S. government securities transfers over FedWire were collateralized. Since uncollateralized daylight overdrafts could fall by 55 percent at all banks (or 66 percent at large banks with $1 billion or more in deposits) as a result of securities collateralization, the estimated rearrangement of federal funds transactions would be correspondingly reduced. If *both* remedies were in operation, daylight overdrafts on FedWire would be largely eliminated. This analysis and conclusion is consistent with one reached earlier using a 1979 sample of FedWire daylight overdraft data.

B. Current Developments in Reducing Interbank Risk on Large-Dollar Payment Networks

Over the past few years there has been a growing recognition of and concern for daylight interbank risk on large dollar funds transfer or payment message networks. In the private banking sector, this concern was recently addressed in three comprehensive reports by the Association of Reserve City Bankers (1982, 1983 and 1984), an overview article by Rosborough and Urkowitz (1983), a brief summary report by bank and thrift institution representatives (FAC/TIAC, 1983), and an issues paper written by the Chicago Clearing House Association (1984). The Federal Reserve System first addressed this issue in a special report to the Board in early 1980, and in four subsequent reports as well. The issue of interbank risk on large-dollar funds transfer networks, and particularly its "solutions," have major implications for:

- The future structure and growth of large-dollar domestic and international interbank payments networks;
- The competitive position among certain classes of banks, such as large domestic correspondents, securities clearers, and foreign banks;
- The use of credit officers and real-time payments and balance automation to access, monitor, and control daylight interbank risk exposures;

- The need to price interbank risk more effectively and recover these risk and automation costs from customers; and
- The evolution of current institutional arrangements in the federal funds and definitive securities markets, and the impact these changes may have on bank use of liability management techniques.

The force and nature of these five implications are dependent upon which daylight overdraft risk reduction alternative is finally implemented. These alternatives and other current developments are discussed next. The implications of these alternatives for bank and financial market operations are discussed in section IX.

Payments Examinations and Access to Net Settlement. As part of the movement from next-day to same-day settlement for CHIPS in October 1981, the Federal Reserve instituted electronic data processing examinations of current CHIPS participants. The purpose of examining participants of this network was to determine the ability of a bank's management to:

1. Assess the inherent credit risks faced in each large-dollar funds transfer network;
2. Establish internal policies and procedures to protect the institution against unreasonable risk exposures and, where deficient, improve these procedures; and
3. Monitor the effectiveness of such safeguards.

Credit risk, as used here, exists when receivers of large-dollar payments allow customers to use these funds prior to final settlement which occurs, except for FedWire, at the end of the day. Current CHIPS payments examinations serve to reduce the risk of loss from:

1. *Credit risk*: by improving internal credit risk controls at banks sending and/or receiving funds over provisional wire transfer networks; and
2. *Systemic risk*: by reducing individual institution credit risk, this also decreases the likelihood of payments system and financial market disruption following a large, unexpected settlement failure on a provisional funds network where one sender failure may lead to others in a systemic manner.

In order to improve further internal controls over credit and systemic risk, the Federal Financial Institutions Examination Council adopted in late 1983 uniform procedures to be used in examinations of wire transfer activities. These uniform examination procedures are based on those currently applied to all CHIPS participants. The FFIEC standardized examination procedures are to be applied in reviews of funds transfer activity for participants on the major payments systems (FedWire, CHIPS, BankWire, CashWire, SWIFT, and Telex/TWX) used by depository institutions.

Payments system examination is closely tied to risk reduction but other risk reduction procedures have been developed. Risk reduction efforts on CashWire were neccessary to achieve access to Federal Reserve net settlement. CashWire already had bilateral net credit limits in place along with 100 percent guarantees by receivers of funds to provide them to customers even if the sending institution failed prior to end of day settlement.[55] CashWire adopted a sender net debit cap of 50 percent of the sender's total capital at the urging of the Federal Reserve as a further risk reduction measure.[56] In the future it is very likely that similar risk reduction procedures will be standard on all large dollar funds transfer networks, either because they are made a condition for access to FR net settlement or for safety and soundness reasons.

Figure 5 represents a schematic picture of a large dollar payment over the three largest wire transfer networks which have immediate or end of day settlement using Federal Reserve reserve accounts.[57] Both the sending and receiving institutions have in place internal controls and procedures for monitoring customer daylight overdrafts or daylight extensions of credit. Such *customer* credit extensions—where funds are sent in excess of customer collected balances or where provisional or uncollected funds received are used prior to settlement—have an associated cost to the customer. However, extensions of *interbank* daylight credit or overdrafts carry no such charges. Interbank overdrafts will receive increasing attention in the future as a result of a movement towards payments finality and the implementation of risk reduction procedures adopted on large dollar networds. The current situation regarding payments finality and risk reduction is now outlined for FedWire, CashWire, and CHIPS.

FedWire: Finality of Payments and Risk Reduction. Reserve Banks, in the event of sender failure, fully guarantee the funds transferred to their receiving bank customers. These receiving institutions, in turn, can treat these funds as collected balances or final funds for their customers. As shown in payments route (1) in figure 5, Reserve Banks, in effect, act as a receiving institution in a FedWire transfer and, because of payments finality on FedWire, face a risk of

---

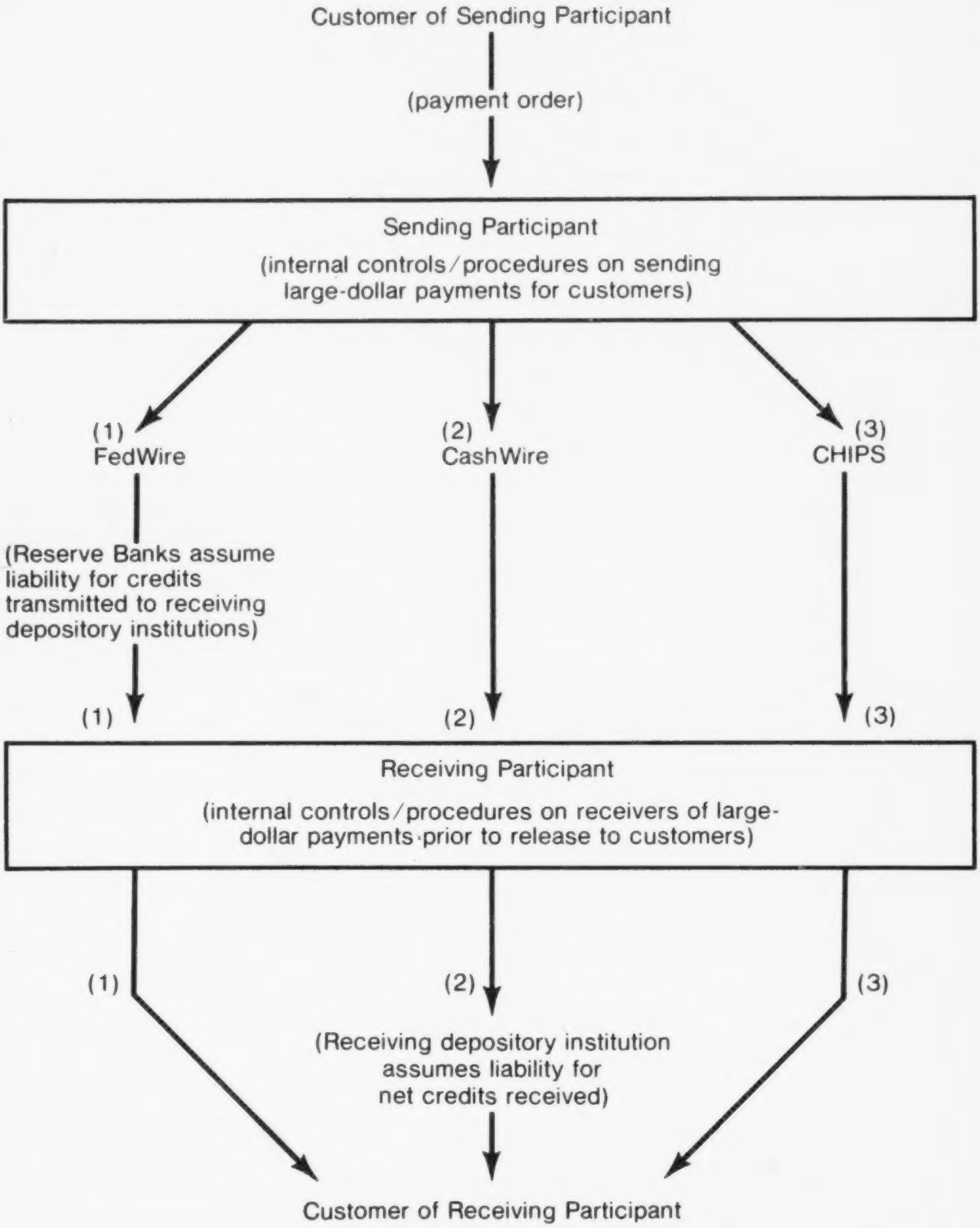
[55] A bilateral net credit limit exists when a receiver sets an internal limit on the value of (net) credits it stands ready to receive (and in the case of CashWire to guarantee) during the day from different senders. It is similar in concept to position limits that banks place on one another regarding the value of federal funds they stand ready to sell each other or the value of CDs they stand ready to buy.

[56] This sender net debit cap limits the sender's net debit or daylight overdraft to 50 percent of its total capital during the day.

[57] The CHESS network in Chicago is dealt with in the text. BankWire and SWIFT, two other "payment networks," are more formally viewed as message networks rather than funds transfer networks. Thus, as noted above, an actual funds transfer need not occur in response to the message sent as balances are merely being transferred between accounts within a single institution. In practice, however, such messages do lead to subsequent funds transfers using FedWire, CHIPS, or some other means.

FIGURE 5

Schematic Representation of Payments Over FedWire, CashWire and CHIPS


Customer of Sending Participant
(payment order)
Sending Participant
(internal controls/procedures on sending large-dollar payments for customers)
(1) FedWire
(2) CashWire
(3) CHIPS
(Reserve Banks assume liability for credits transmitted to receiving depository institutions)
(1)
(2)
(3)
Receiving Participant
(internal controls/procedures on receivers of large-dollar payments prior to release to customers)
(1)
(2)
(3)
(Receiving depository institution assumes liability for net credits received)
Customer of Receiving Participant

loss for which no compensating charge is now assessed.[58] There are two procedures which should assist Reserve Banks in reducing this risk, however.

For FedWire funds transfers, Reserve Banks now counsel a sending institution whose daylight overdraft to capital ratio exceeds certain guidelines. In addition, daylight overdrafts for Edges, and branches and agencies of foreign banks are required to be fully collateralized. The daylight overdraft to capital ratio guidelines act in effect as flexible bilateral net credit lines, limiting the net credits extended by Reserve Banks to senders in daylight overdraft. Instead of varying these Reserve Bank determined bilateral net credit limits on the basis of an explicit credit judgement, these limits are related to a sender's capital position and in principle uniformly applied across all sending banks.

A different treatment is to be adopted for FedWire securities transfers. Here the U.S. government book-entry securities being transferred are to be used as collateral for daylight overdrafts resulting from securities transfers. Daylight overdrafts associated with transfers of government book-entry securities account for over 55 percent of the value of FedWire daylight overdrafts experienced for all banks (66 percent for large banks with $1 billion or more in deposits). The final legal and operational details of this procedure are still being developed. The underlying rationale for treating securities transfers differently from funds transfers is to minimize possible disruption in the government securities market because of the important role it now plays in the implementation of monetary policy through open market operations.

CashWire: Finality of Payments and Risk Reduction. On CashWire, receiving banks guarantee funds transferred in the event of sender failure. This is shown as route (2) in figure 5. Customers can treat funds so received as collected balances. Thus on CashWire, like on FedWire, there are two guarantees for each payment transaction—one from the sender, the other from the receiver. One difference, of course, is that Reserve Banks (in principle) have unlimited access to settlement funds in the event of sender failure whereas receiving banks on CashWire do not. Another difference is that CashWire is new and still small compared to CHIPS. Consequently the apparent ease by which receiver guarantees were implemented on CashWire can not be used as an accurate precedent for their imposition on an already established and very large system such as CHIPS.

The receiver guarantees on CashWire are not open-ended.

---

[58] In the chain of events whereby Bank A sends a payment to Bank B over FedWire, Reserve Banks are viewed as the first receiving institution or the second sender. In the former case, FedWire provides a receiver guarantee; in the latter case it provides a sender guarantee.

Bilateral net credit limits are in place and these limits set a ceiling on the net dollar amount a receiving institution stands ready to accept and pass on as "final" funds to its customers. In this manner, bilateral net credit limits reduce payments system credit risk, and therefore systemic risk as well. As an additional risk reduction measure, CashWire senders have a limit placed on the net debit they can incur as a percent of their capital, as explained above. There is no use of collateral at this time on CashWire.

CHIPS: Finality of Payments and Risk Reduction. On CHIPS, there is no receiver guarantee in the event of sender failure (see route (3) in figure 5). Instead, sender failure can lead to an "unwinding" of the failed sender(s) payments and receipts from all transactions made and the computation of a new settlement statement. Receiving institutions and their customers should treat funds received over CHIPS as provisional, rather than final, as over FedWire or CashWire. There is information which suggests that this distinction is not always made as clear to the customer as it could be. While there are no receiver guarantees on CHIPS today, bilateral net credit limits are being tested and are to be in effect during 1984. At present, there are no sender net debit limits or use of collateral on CHIPS. However, there has been a decision to establish sender net debit limits but their level has not yet been determined. Finality of payment is currently being studied.

CHESS: Finality of Payments and Risk Reduction. The Chicago Clearing House Association rules regarding the (original) five Chicago bank CHESS network state that "All payment messages shall be treated as immediately available funds upon receipt." However, these rules do not provide for any receiver guarantees or collateralization to ensure that this is the case. In the event of sender failure, an "unwind" occurs and a new settlement statement is to be computed, like on CHIPS. At the present time, CHESS has bilateral net credit limits and a 50 percent of capital sender net debit limit in place. There is no use of collateral on CHESS and finality of payment seemingly does not apply.

C. Future Developments in Reducing Interbank Risk

Likely Future Developments. Given the current concern of the Federal Reserve and certain banking sector groups for reducing risk on large dollar payments networks, and considering the direction that current efforts are proceeding, it is very likely that new risk reduction procedures will be adopted in the near future. The most probable outcome would appear to be:

(1) The adoption of *some form of receiver guarantee*, to be used in conjunction with bilateral net credit limits, by all current and new large-dollar networks; and

(2) The use of sender net debit "caps," set relative to a sender's capital, which apply to funds sent across all large-dollar networks. To provide flexibility, some form of daylight "overline" credit may be provided at a fee by banks or the Federal Reserve.

As seen from the preceeding discussion of payments finality and risk reduction procedures on the three largest large-dollar networks, elements of these two likely future developments are consistent with those already fully or partially in place today. The main sticking point, aside from the determination of an appropriate phase-in schedule, will likely be whether or not the receiver guarantee should be 100 percent or some smaller number; and, whatever the number chosen, whether or not receivers making the guarantee should share this increased risk with senders by having senders collateralize all or part of their net debit. This point is quite important since bilateral net credit limits with receiver guarantees means that the receiving bank will no longer have recourse to its customer for funds credited to the customer's account in the event of sender failure. This increases the receiving banks' risk of interest and principal loss and, the argument goes, provides a strong inducement to set prudent bilateral net credit limits to keep this risk within manageable bounds. This, in turn, reduces the probability that sender failure will lead to settlement failure, thus reducing systemic risk. But the reduction in systemic risk (if realized) comes about by increasing, at least initially, individual bank credit risk. This is because receiving banks may be more confident of recouping provisional funds used by their customers (when they have recourse to their customer) than in recovering net credits from a failed sending bank when receiver guarantees are in place (and there is no recourse to the customer).

Making Receiver Guarantees Acceptable On CHIPS. Some possible solutions to this very real and valid concern regarding receiver guarantees on the part of receiving banks would be to:

(1) Use partial rather than 100 percent receiver guarantees (e.g., where \$.50 of each dollar lost on net credits received within the bilateral net credit limit is guaranteed and paid by the receiving bank; the remaining \$.50 of each dollar of net credit loss is recouped by recourse to the receiving bank's customer);[59]

(2) Treat net credits received within the bilateral net credit limit as

---

[59] The case today on CHIPS is that 100 percent of the loss on net credits from a failed sender can be recouped by recourse to the receiving bank's customer, to the extent provisional funds are used prior to settlement and to the degree that the customer is both willing and able to return these funds. The customer, in turn has legal recourse to the failed sender and the sending customer of the failed bank.

final funds—through 100 percent receiver guarantees—but treat net credits received over the limit as provisional funds (with recourse to the customer);[60] or

(3) Adopt 100 percent receiver guarantees but require that senders collateralize all or part of their net debit. This increases the probability that receiving banks can recover net credits from a failed sending bank (using the posted collateral if necessary).

While any one or mixture of these arrangements would not please everyone, two important results would be attained. First, there would still be at least some increase in credit risk to receiving banks, ensuring that the bilateral credit limits would be more prudently set than if receiving banks faced no risk of loss at all. And second, while some risk would be absorbed by receiving banks, this risk of loss would be much less than the daily billion dollar daylight exposures that now exist on CHIPS, exposures which at times exceed a bank's total capital position. Given the size of these CHIPS daylight exposures, it is clear that they can not all be absorbed by receiving banks without significantly disrupting current payments operations. For this reason, the risk of sender failure needs to be shared with a receiving bank's customer (as in option 1 or 2 above) or shared with sending banks (as in option 3). Such a conclusion also follows from examining some general principles of risk assignment.

Three Principles of Cost/Risk Assignment. The payments finality and risk reduction alternatives discussed above typically impose costs and/or risks on either only the receiver (as in bilateral net credit limits) or only the sender (as with sender collateralization or net debit "caps"). Although depository institutions are usually both senders and receivers of large dollar payments, the mix is neither equal nor stable across institutions. The question then is who—receivers, senders, or both—should bear the costs and/or absorb the risks involved in reducing systemic risk on large-dollar payments networks.

There are three generally accepted principles guiding the economic assignment of liability and or costs in such an arrangement:[61]

(1) Liability should be borne by the party in the best position to absorb, spread, and/or insure against the cost or risk of settlement failures;

(2) Liability should be borne by the party in the best position to detect and prevent such failures; and,

---

[60] In order not to perpetuate the current situation, some mechanism would have to be put in place so that the bilteral net credit limit could not always be set at zero, making all net credits received provisional to all customers (which is the situation on CHIPS today). One possibility would be to permit customers access to funds only if they were final, not provisional. Today, of course, customers have access to provisional funds prior to settlement.

[61] Another principle of risk sharing could be that all participants in a payments network share *pro rata* in any losses accruing to the network. This arrangement, however, does not provide proper incentives for risk reduction because creators of risk are not held fully responsible for the risks they create. This is the so-called moral hazard issue, one which is well-known in the insurance business. A settlement insurance alternative is discussed below.

(3) Liability should be borne and procedures adopted by the party being relied upon to provide "good funds" for the payments sent to ensure that settlement can be promptly completed.

On a large-dollar payments network, these three principles would perhaps lead to the following assignments of liabilities:

- The receiver, by allowing customers to use payments credits prior to end of day settlement, creates systematic risk if the sender fails. The receiver's customers benefit by using payments credits prior to settlement and so provide a justification for and a mechanism which could be used to charge for or insure against the risks created—namely, charge the customers receiving and using credits prior to end of day settlement;
- The receiver appears to be in the best position to assess the credit risk involved in receiving payments from different senders *and* the credit risk created by allowing customers to use (all or a portion of) the credits received prior to end of day settlement; and
- The sender is the party being primarily relied upon to provide "good funds" for the end of day settlement. His failure starts a sequence of events which creates systematic risk. Such risk would be reduced or avoided if senders were required to post collateral for their daylight net debit positions. Alternatively, and/or in addition, such net debit positions could be subject to an upper limit or "cap."

On the basis of these three principles, it would appear that perhaps the receiver alone, or more likely the receiver and the sender, but not the sender alone, should appropriately initially be assigned the costs and/or risks of reducing the credit risk and systematic risk of settlement failure.[62]

Sender Net Debit Caps. The sender restriction most often discussed is a sender net debit cap. This sender restriction, depending on how it is set, may or may not be more restrictive than if current loan limit rules were extended to cover intraday customer and interbank exposures as well as longer term loans. The Comptroller of the Currency in 1982 put such a proposal out for comment but, in the final revision of the loan limit rule, decided not to extend them to intraday exposures at this time. The similarity between loan limits and sender net debit caps suggests that the same sort of examining procedures could be used for enforcement if caps were adopted. No expensive and operationally complex real-time monitoring of caps would be needed across all large dollar networks. Such monitoring and control could be done by the bank itself as occurs for loan limits. Each network individually, however, could monitor and control the exposure of each of its participants but the sender could be made responsible for the total exposure across the various networks.

---

[62] While charges may be imposed on customers receiving credits, these charges may, through the workings of the market, be shifted to others.

Another important consideration regarding sender net debit caps concerns the different funding procedures for similarly sized banks in unit banking *vs*. branch banking states. A large bank in a unit banking state will, on average, fund a larger percentage of its total assets by purchasing a smaller bank's deposit funds than will a large bank in a branch banking state. In the later case, smaller banks can, in effect, be part of a single branch banking organization contributing deposit funds to the head office through *internal* funds transfers. In unit banking states, large institutions purchase such funds using *external* wire transfers and so would likely find a sender net debit cap (when repaying borrowed funds—usually each day) more restrictive than a similarly sized branch banking organization. Equity would seem to require that sender caps, if implemented, reflect these differing funding needs (on average) due to current differences in state branching, and therefore funding, restrictions.[63] An alternative view would be that equity is second in importance to actual risk reduction and that such reductions should be uniform accross banking organizations regardless of the state in which they happen to be located.

Lastly, depending on the level of the cap adopted, it may be important to allow for some flexibility in its implementation in order not to unnecessarily disrupt the operation of payments systems and financial markets. If the cap adopted would be binding on some or most institutions today, it would be necessary to carefully design a phase-in schedule over a period long enough to allow for nondisruptive incremental changes in payments practices to occur. Alternatively, flexibility could be achieved by having a group of banks or the Federal Reserve provide daylight "overline" credit at a fee. Such an arrangement would permit sending institutions to stay within their cap even when making certain exceptionally large payments, as when one bank acts as a disbursing agent (or concentration bank) for a very large loan takedown in which many banks have participated. Of course, charging a fee for daylight credit provided above a cap will likely induce banks to seek out more cost effective ways to reduce the incidence of daylight overdrafts. These are discussed below in section IX.C and, as will become obvious, there are many different things banks can do to reduce daylight overdrafts short of creating a payments "gridlock" that some have argued would be associated with the implementation of sender caps.

---

[63] Assuming that all net purchased federal funds reported by unit and branch banking organizations are made externally using FedWire, unit banks with over $2 billion in deposits purchase (on average) some $160 million per bank more (net) federal funds than does a branching bank. In terms of a sender net debit cap set as a ratio to total capital, this implies that such a cap for unit banks would have to exceed a similar cap for branching banks by an average of .85 to fully compensate unit banks for the inability to have an extensive branching network to raise funds internally. That is, if a branching bank's sender net debit cap to capital ratio is 2.0, the "equitable" cap for the average unit bank would be 2.85. The extent to which some net federal funds purchases are book-entry, and therefore internal to a bank since FedWire is not used, can not be determined without a special survey.

D. Other Alternatives for Reducing Interbank Risk (and Reasons for Their Disfavor)

Some other alternatives which have been suggested in earlier risk reduction discussions are:

(1) Institute rolling settlement, whereby a large-dollar network settles more than once each day;
(2) Purchasing settlement insurance from a private company;
(3) Develop indemnification agreements between network participants; or
(4) Prearrange a collateralized loan and charge a commitment fee for the credit line extended.

These alternatives are now briefly discussed and reasons for their current rejection by one group or another are noted.

Rolling Settlement. Bilateral net credit limits and sender net debit caps have one flaw in common: they encourage senders of large-dollar payments to send later in the day. If only one or a few senders send later in the day, the net result will be that these senders' daylight overdrafts will fall. This reduces the probability that these sender's net debit caps will be reached and increases the likelihood that their credits sent will also be within the bilateral net credit limits set by receivers. But any improvement for these senders results in increased daylight exposures of all other senders, reversing the two beneficial effects just noted. As a result, other senders will also attempt to delay payments until later in the day. If all senders attempt to delay payments, then no sender may gain. Just as important, any sender who finances his intraday exposure with overnight or (to be developed) intraday funds risks passing on to his customer financing costs which may not be incurred by senders who continue to delay payments rather than fund their intraday exposures.

In sum, senders are given a strong incentive to delay payments to (a) shift daylight exposure to other senders or (b), if this exposure is covered by intraday funding, shifting the cost of this funding to other senders. Rolling settlement has been suggested as one means whereby incentives to delay payments would be reduced. If a sender delayed payments until much later in the day, he may succeed in reducing daylight exposure in the morning prior to, say, noon settlement but only at the cost of increasing his afternoon exposure prior to end of day settlement.

Rolling settlement, however, has been basically rejected because it could prove operationally difficult to implement and because a simple "rule," if followed, could generate the same results. Such a rule could stipulate that, for example, no more than 40 percent of each day's payments (or the day's payments made in the morning) should on average be sent in the afternoon. A rule like this limits the ability of senders to shift their daylight exposure to others and, importantly, would likely be operationally easier to implement than rolling settlement. In addition, it would provide an inducement for sellers of overnight federal funds to send funds earlier in the business day than now occurs. This result would reduce the time, and consequent daylight exposure, between the traditional a.m. return of overnight federal funds and the p.m. receipt of repurchased funds for the next night. If realized, this could reduce the daylight exposure due solely to the institutionalized operation of the federal funds market.

Settlement Insurance. If the potential dollar losses due to settlement failure could be expressed in terms of sound actuarial principles, it might be possible to find a group of large insurance firms willing to provide settlement insurance. The problem is that settlement risk is not actuarial in nature. One can not even approximate the probability of settlement failure on CHIPS or any other large network by taking the ratio of settlement failures to total settlements—the frequency approach to probability. Such a ratio is zero as no settlement failure has yet been observed. What is needed is a subjective but expert guess of the probability of settlement failure—a Bayesian approach to probability. To date the insurance industry has been unwilling to make such a subjective judgement and determine the probability that settlement losses, which could run into billions of dollars, could be incurred. In short, because the risks are difficult to quantify and the potential losses are so large, the insurance industry representatives contacted by Federal Reserve staff were not interested in providing settlement insurance, even if the risk could be diversified across many large insurance firms. However, if such an insurance policy were to be issued, its price could perhaps be based on the spread between the rates paid on 90-day commercial bank large-denomination certificates of deposit (or 90-day commercial paper issued by bank holding companies) and 3-month Treasury bills, adjusting for different tax treatment. This would be a measure of the market price of default risk for settlement insurance for FedWire. This spread, which has over 1983 averaged around 150 basis points, could in principle serve as a useful guide for determining the premium rate that should be paid on a prespecified amount of insurance coverage. Applied to an insurance limit of $1 billion, this premium rate would be quite expensive; it would cost some $15 million a year. Even use of a "dayloan" rate of 100 basis points results in very significant costs to banks which incur daylight overdrafts. But then, these costs provide the right indentives; namely to reduce use of currently "free" daylight credit, creating risk which is not compensated for.

Indemnification Agreements. An indirect method to price the risk involved in net settlement would be for all participants in a particular funds transfer network to agree to share (on some *pro rata* basis, such as average dollar volume) the loss resulting from a participant's failure to settle in a timely manner. These indemnified losses would include any loss resulting from a single or group of net debtor participants having insufficient funds to cover the full amount of its settlement debit. An unanswered question concerns the appropriate accounting treatment of such an indemnity. Participants may be required to classify such a settlement indemnity (limited or open-ended) as a contingent liability; it is probably not equivalent to a standby letter of credit. The obvious shortcoming of indemnification is that surviving participants may have insufficient resources to cover the losses generated by settlement failures. Furthermore, in the absence of coordination between the participants, indemnification distorts incentives for risk control. The adverse consequences of a settlement failure are borne by parties not responsible for the failure. This discourages any participant from prudently managing risk as long as at least one other participant's behavior is lax—the moral hazard issue.

Prearranged Collateralized Loan with a Commitment Fee. A prearranged loan either at the Discount Window or at participant banks could be structured to virtually eliminate any liquidity problem of a participant institution and so insulate large-dollar networks from settlement failure. On the negative side it has been argued that such an arrangement, if tied to the Discount Window, would provide a degree of automatic recourse to Federal Reserve credit—and hence reserve provision—that may be undesirable from a monetary control point of view and possibly encourage risky behavior.[64] Perhaps more importantly, such an arrangement could set a precedent for other types of "guaranteed access" to the Window and thereby reduce its usefulness as a separate monetary policy instrument or credit policy tool in the future.

Other considerations have been raised regarding this alternative. These are: (1) that CHIPS daylight exposures are so large that loan limits to a single borrower could come into play if the loans were made by commercial banks (rather than Reserve Banks with access to the U.S. Treasury); (2) that defining and maintaining proper collateral would not be easy; and (3), that loss of collateral increases the risk exposure of the FDIC and uninsured and unsecured creditors in the event of sender failure. The last argument may be circumvented if the FDIC and uninsured/unsecured creditors charged higher fees to more risky banks. For example, if the FDIC adopts risk-related deposit insurance fees, one determination of this fee could be the amount and frequency of daylight overdrafts. If overdrafts greater than some predetermined value continue to occur, this information could be published in the bank's quarterly report to the SEC, so that uninsured depositors and debt holders would be informed of the risk which exists and perform a market regulation function (e.g., demanding a higher risk premium to rollover funds when they mature). This approach, however, may not be altogether effective because the likelihood of a deposit assumption (where uninsured creditor losses are low or zero), rather than a deposit payout (where these losses are larger), is very high under current FDIC procedures if a large sending bank failed. Thus market regulation may be weak in practice. And under current FDIC proposals for risk-related deposit insurance premiums, the range of premium variation is relatively small and may not be large enough to significantly alter bank behavior (although this remains to be seen). For these reasons, this alternative has not been favorably received at the Federal Reserve even though it reflects how bankers have traditionally dealt with the credit risk issue; namely use collateral for a loan and charge higher risk premiums for more risky borrowers.

---

[64] A counter to this argument is that: (a) such credit would rarely be needed so the negative monetary control effects would be virtually nonexistent in practice; and (b), charging a commitment fee related to the size of the daylight exposure, and hence the size of the prearranged loan, should provide a proper inducement to prudently limit risk even if the loan itself were assured beforehand.

E. Examples of Risk Management Arrangements Outside of Banking

Certain risk management arrangements are already operational on commodity exchanges. Position limits, collateralization, and indemnification are commonly used by stock and commodity exchange clearinghouses to ensure the integrity and final settlement of transactions. The simplest position limit sets a ceiling on a clearinghouse member's net interest on a contract. A member's net interest on a contract represents the difference between the value of the outstanding contracts in which the member has taken a long position and the value of the oustanding contracts in which the member has taken a short position.[65] A member with a net long position incurs losses when the commodity's price declines, while a member with a net short position incurs losses when the commodity's price rises. (A member whose net position in neither long nor short does not gain or lose when the commodity's price fluctuates.) Since there are limtis on daily price fluctuations, by setting a ceiling on a member's net interest the size of the loss which a member can incur (the clearinghouse's potential liability) is bound.[66]

---

[65] Net interest on a commodity futures contract is analogous to a commercial bank's net position in a foreign currency.

Although their collateral requirements vary, every clearinghouse requires each of its members to deposit a specified amount in a special guaranty fund. These guaranty deposits range from as much as $2 million for each member of the Comex Clearing Association having net capital of $20 million or more to as low as $7,500 on the New York Cotton Exchange. In addition, all of the clearinghouses have specific procedures to offset the losses of a member defaulting on its obligations. Generally, the defaulting member's obligations are satisfied from three sources of funds: (1) the assets of the defaulting member under the control of the exchange (its open contracts, its margin deposits, and its contributions to the exchange's guaranty fund), (2) a reserve representing an allocation of the exchange's net worth, and (3) the guaranty fund deposits of all the surviving members.[67] If these three sources prove to be insufficient to satisfy the failing member's debts, the remaining obligations are paid off by assessing each surviving member. Each member is assessed either an equal amount or a *pro rata* share based on factors such as its trading volume, open position, financial condition and the types of contracts that it trades.

For completeness, it is worth noting that other industrialized nations' central banks do not provide daylight overdraft credit, other than that given through a discount window facility for periods of one day or longer. If discount window credit is given, it is fully collateralized and (typically) a fee is levied. One reason for this different treatment of daylight overdrafts is that the banking system in other countries is much more concentrated than that in the U.S. Two results follow from this institutional difference:

(1) Banking concentration reduces the volume of wire transfers needed between banks, since more large dollar transfers are likely to be handled internally between accounts within a single banking organization; and

(2) No other central bank provides a wire transfer network like the Federal Reserve FedWire system, although net settlement is provided.

---

[66] Within the net settlement context, the transfer network is analogous to the clearinghouse and position limits would represent ceilings on a network user's net credit and debit positions.

[67] Within the wire transfer context, the margin deposit is somewhat analogous to a correspondent balance or a balance maintained at a Reserve Bank. The difference is that unlike the balances required by correspondents and the Federal Reserve, the required margin deposit fluctuates on a daily basis in response to changes in the risk exposure of a commodity clearinghouse member.

The first result means that the incidence of daylight overdrafts in other countries would, if they occurred, be much lower merely as a consequence of banking concentration. The second result implies that Federal Reserve participation in the payments system in general, and wire transfers in particular, is due primarily to the historical lack of banking concentration in the U.S. Together these results suggest that the increase in U.S. banking concentration which could follow intra and interstate banking will reduce daylight overdrafts, by reducing the need for externally provided wire transfers, and thereby reduce the demand for FedWire services. This line of reasoning is more fully discussed below.

IX. The Impact of Costs, Competition, and Risk in Changing the Payments Mechanism

Three major elements which will determine the shape of the future payments mechanism concern: (a) the relative costs of the different payment services; (b) the current competitive structure and degree of future banking concentration; and (c) how the risk issue on large dollar payment networks is resolved. Each of these three determinates and their likely effect on the payments mechanism are discussed in this section. The specific issues addressed concern:

- Costs—paper *vs.* electronic based processing methods;
- Competition—national *vs.* locally supplied payment services and the substitution of internal for external payments; and
- Risk—provisional *vs.* final settlement of large-dollar payments.

This discussion relies, in part, on analyses and empirical results presented earlier.

A. Costs: Paper vs. Electronics

Predictions of electronic payments replacing paper based payments have almost always been exaggerated. This is because the focus usually has been on the technological feasibility of using electronics to replace checks. But while technological feasibility is clearly a necessary condition, it is not sufficient. Cost savings have to be demonstrated and, after that, users have to be convinced that the cost savings outweigh the benefits of continued reliance on existing and satisfactory payment methods. Public acceptance of new electronic payment methods has always lagged behind their technological feasibility. This argument was clearly (and convincingly) set out in Cox and Metzker (1983).

Aside from a small segment of users, the substitution of electronic payment methods (ACH, automated teller machines, cash dispensers) for checks has not been rapid. Generally, such substitution as has occurred has been due to:

- Increased convenience (as in the use of automated teller machines and cash dispensers which are more accessible either by place or time of day or both);
- Float gain (as in receiving—but not sending—ACH payments); and

- Reduced costs (as in bank substitution of cash dispensers for teller and branching costs or accounting/postage cost savings in exchanging corporate ACH trade payments).

Continued limitations in the shift to electronics exist, however, and concern: loss of check float for ACH payors; the safety and privacy of the payment information; legal liabilities for losses and fraud; and simple inertia in switching to a newer payment method without significant additional monetary incentives and/or a basic dissatisfaction with existing arrangements. With these broad based, but largely nonquantifiable, considerations in mind, the implications for payments product mix of the above noted scale economy and relative processing cost differences are now analyzed for wholesale payments services.

ACH as a Substitute for Check Processing. In 1983, the average production cost (excluding overhead) of automated clearing houses per processed electronic funds transfer image was 4.7 cents. For paper checks processed by the Federal Reserve the production cost was only 1.4 cents. ACH costs would have to fall by 70 percent to match those of checks, neglecting the cost of Federal Reserve float. Given the 1983 average volume per ACH office of 13 million, at what higher volume could ACH average production costs match those of paper checks? Holding all else constant and changing only ACH volume in the ACH statistical model estimated above, an illustrative calculation suggests that ACH volume at the average office would have to rise to nearly 88 million images per year to match paper-check costs.[68] The same result is obtained even when overhead costs are included. The result holds because the ratio of overhead to total costs for checks is almost the same as that for ACH processing, so that equality of production costs implies equality of total unit costs.

Although a volume of 88 million images per office substantially exceeds current ACH levels, it should be obtainable because it is one-third of the average volume of processed checks at Federal Reserve offices (270 million). Although merely illustrative of what the future for ACH costs may be, these calculations provide some empirical support for the conclusion that ACH processing may be a viable and potentially lower-cost substitute for paper checks in processing funds transfers at the wholesale level. Processing costs at the wholesale level, however, represent only a small part of the cost of making an ACH or a check funds transfer. Equality of processing costs, therefore, does not imply equality of the full cost of a funds transfer to the final consumer, as seen in table 2. The issue here is that while some intermediate users may at some point face incentives to use ACH over checks, final users (consumers) are currently a long way from facing the same incentives.

---

[68] For the 24 Federal Reserve Branch ACH offices, the statistically estimated cost model was $\ln TC = 7.16 + .603 \ln Q$ using 1980, 1981, and 1982 data in a pooled regression (see the appendix for a further discussion of the statistical model used). The time period shift variable was $.24 \ln t$, where t is 1, 2, and 3 for the three years but is neglected in the extrapolation to accurately reflect only the scale economy effect. The problem is to find the Q at which $\exp(\ln TC)/Q = \$14.00$ = average cost per 1,000 images (Q is measured in thousands of images processed: $14.00 per thousand equals 1.4 cents per image). Substituting the estimated equation into the formula yields $Q = 88$ million images. These estimates are, however, very sensitive and should only be considered as first approximations or illustrations.

B. Competition: National vs. Local Services

Retail Payments Services. Competition in the provision of retail payments services is basically local and could remain so even with interstate banking. Local banks should be able to compete for two reasons:

(1) Franchise arrangements with larger banks will give them the ability to offer a broad line of payments services; and

(2) Large banks seemingly have no real cost advantage (and may face actual scale diseconomies) in providing banking services so smaller banks can be price competitive.

Franchise arrangements already exist with small banks serving as distribution points for money market deposit accounts, trust services, and even international payments products. Franchising national retail payments services will become an important correspondent bank service, just as is now the case for the wholesale services of check processing, wire transfers, loan participations, and (to some limited degree) interstate automated teller machine access.

The commercial bank average cost and scale economy results discussed above clearly support the contention that larger, nationwide banks need not have a marked cost advantage over smaller banks in providing basic banking services. However, small bank profitability may fall as increased competition erodes their ability to earn higher than average revenues on loans and pay lower than average interest on deposits. Thus the competitive issue in interstate banking may not be one of an inability of small banks to have access to and supply retail payments services at competitive costs as much as it is a concern that increased competition will make the demand for their services more elastic and reduce their revenues. They will be able to compete, but their return on equity should be reduced.

Wholesale Payments Services. Competition in providing checks, ACH, and wire transfer wholesale payments services has a short-term and a longer-term dimension. The short-term dimension has already been extensively discussed, focusing on new entry into check and ACH processing and the likely emergence of new local/regional large-dollar wire transfer networks. The longer-term dimension concerns the effect that interstate banking and the development of a reasonably concentrated national retail payments service market can have on the market for wholesale payments services.

In the long-run, the role of the Federal Reserve in all types of payments processing should fall both relatively and absolutely. The trends leading to this result are already underway and concern:

(1) Demand for payments processing services (either internally or externally supplied) will increase at a decreasing rate in the future as the growth in retail transactions volume increases at a decreasing rate;
(2) Movement to intra and interstate banking will shift an increasing share of payments processing from external to internal sources and reduce the Reserve Bank share of a (slower growing) payments processing market; and
(3) Additional Reserve Bank share reductions will be associated with future developments regarding check truncation and increased use of automated teller machines, cash dispensers, and point of sale systems, all of which shift retail payments from checks to electronics.

In simple terms, (1) implies that the payments processing "pie" will not grow as fast in the future as it has in the past while (2) and (3) suggest that the Reserve Bank share of this "pie" will be reduced over time from interstate banking and the slow but continued shift away from checks. Short-run developments or options within Federal Reserve control can clearly speed up or slow down this longer-run shrinkage of Reserve Bank operations but, in our judgement, cannot reverse the decline.[69]

Future Growth in Transactions Volume. The likelihood that growth in transactions volume of all types will increase at a decreasing rate in the future is based upon projections that the growth in household formation in the 1980s and 1990s will be less (percentagewise) than that experienced in the 1960s and 1970s. The post World War II baby boom cohort group has reached adulthood with family/household formation in 1960s and 1970s but no large age distribution distortion will occur until the turn of the century, when the children of baby boom adults reach adulthood themselves. As a result, the future growth in the payments processing "pie" will be markedly slower. This slower growth in transactions volume implies a slower

---

[69] Some short-run developments which would speed up or slow down this shrinkage are:

Speed Up:	Slow Down:
Offer a DSS service.	Offer an acceptable ECC service.
Offer a separately priced check and ACH return item service.	Offer international FedWire services.
Encourage check truncation (so fewer checks need to be collected physically).	Reduce the growth and level of ACH and wire transfer processing costs.
	Offer payments/accounting services directly to nondepository institutions.

growth in the derived demand to process payments transactions at the wholesale level.

Effects of Interstate Banking. A movement to intra and interstate banking will permanently reduce the volume of check, ACH, or wire transfers requiring external—Federal Reserve Bank, correspondent bank, or independent service bureau—processing. This is because interstate banking, by increasing the size and geographical dispersion of banking organizations, will increase banking concentration and result in larger volumes of "on us" payments. "On us" payments only require the internal movement of funds between accounts at the same banking organization rather the external movement of funds between accounts at different banks. Greater banking concentration will also lead to a consolidation of check presentment endpoints making direct presentments by banks more cost-effective. In turn, these developments should increase incentives for direct exchanges between local/regional check clearinghouses and originators and receivers of ACH and wire transfers.

The logical conclusion of increasing banking consolidation is seen in Canada, England, and Europe where only a handful of very large banks exist. Here these nations' central banks do not supply any payments processing services (save for net settlement) and all payments transactions are handled internally or at a very large national payments clearinghouse.[70] While U.S. banking concentration may never equal that in other nations, it could rise rapidly with interstate banking. One recent survey projects that the number of separate banking organizations may fall by one third by 1990. A second analysis suggests that if all bank mergers that could occur under present Justice Department guidelines (which use a Herfindahl Index) *did* occur, less than 20 banks may exist nationwide as local markets are consolidated.[71]

In the past, bank mergers have been preferred to *de novo* entry—establishing new branch offices—as a means to enter new markets. In fact, Rhoades (forthcoming) has shown that mergers have accounted for 72 percent of the current size of the 20 largest U.S. banking organizations. This suggests that when interstate banking occurs, its effects on reducing the share of Reserve Bank payments processing volumes is more likely to be rapid than slow, since internal growth is more rapid through mergers than through *de novo* entry into new geographical markets. In addition, as has been seen in the flurry of regional interstate banking activity, interstate banking through mergers between large banks is more likely than mergers between small banks (or between large and small banks). As a result, the highest payment volume endpoints will likely merge first and more rapidly reduce the need for external Reserve Bank payments processing than if mergers were concentrated between large and small or only small banks.

---

[70] For example, in Canada the 6 largest banks hold 94 percent of total assets—$350 billion in 1981—and the Canadian central bank provides no payments services except settlement.

[71] These studies are, respectively, an Arthur Anderson & Co. and Bank Administration Instititue judgemental survey summarized in the *American Banker*, June 22, 1983, p. 1, and Burke (1984). The Burke study illustrates just one of the many ways in which banking concentration could rise, namely through concentration in separate local markets so the nation as a whole looks like a large local market with roughly equal participation in each local area.

Reduced Use of Checks. Future developments in check truncation, use of automated teller (ATM) networks and cash dispensers (CD), and establishment of point of sale systems (POS) all point to the continued (slow) replacement of checks by electronics in retail payments (c.f., Cox and Metzker, 1983) As noted in earlier discussions of new entry into wholesale payments processing services, Reserve Banks are not likely to expand check market share and can only lose ACH share. Thus the transition from paper to electronics should reduce the future demand for Reserve Bank payment services compared to what it would otherwise be. In addition, previous Federal Reserve policy statements have suggested that Reserve Banks will not play a role (other than net settlement) in providing processing services to retail ATM, CD, or POS systems but, instead, will continue to specialize in wholesale payments services. Thus the substitution of these electronic payment methods for checks should contribute to the slower growth, if not reduction, in the derived demand for Reserve Bank check processing services.[72]

C. Risk: Provisional vs. Final Settlement Systems

Among the more controversial ways to reduce the credit and systemic risk associated with large-dollar wire transfer networks would be to adopt payments finality. With payments finality, or receiver guarantees as discussed above, once a wire transfer is received at a bank, the receiving customer would have "good funds" or collected balances credited to his account even if the sender failed to settle. Such a system is in operation on FedWire and CashWire but not on CHIPS or CHESS. There are at least four important implications of moving from provisionality to finality on large-dollar networks:

(1) Changes in the future structure and growth of large-dollar networks;

(2) Shifts in the competitive position between banks, depending on the type of payments business they specialize in;

---

[72] But even if Reserve Banks did play a major role in both retail and wholesale electronic payment services, it would still imply a shrinkage in the number of Reserve Bank offices. Electronic payment services are less labor intensive and typically more cost-effective when centralized, unlike checks which require decentralized operations and are labor intensive.

(3) Changes in the way customer accounts are monitored, controlled and priced (to recover the automation and risk costs associated with finality); and

(4) Changes in the institutional structure of financial markets.

These implications are now discussed.

Future Structure of Large-Dollar Networks. Access to the Federal Reserve's net settlement service has been generalized and made more equitable. This should encourage the future establishment and growth of new local/regional large dollar funds transfer networks. Other incentives have been the establishment of prices for FedWire and a desire in certain banking circles to develop alternatives to Federal Reserve (FedWire) and large New York banks' (CHIPS) controlled networks. Entry of new suppliers will decrease the influence that the Federal Reserve and large New York banks now have over how these services are developed in the future. The range of services which can be offered over these networks will thus likely be expanded and thereby affect the competitive position of the different banks using different networks.

Competitive Position Between Banks. Implementation of bilateral net credit limits and sender net debit caps, if realized, can significantly affect the competitive position among different classes of banks, such as large domestic correspondents, securities clearers, and foreign banks. This is because if net credit limits and sender caps are tied primarily to bank size or capital levels, then some senders may find themselves unable to maintain the payments volume they now have[73] Without some procedure which would permit these limits to be consistently exceeded, such as collateralizing sender net debits in excess of a cap or permitting end of day receipt *after* settlement occurs for those funds received in excess of a bilateral net credit limit (i.e., option 2 in section VIII.C), some correspondent banks with very large payment volumes will lose market share. As a result, past investments in payments automation may not earn the returns forecasted for them when no risk reduction programs were in place.

Large U.S. government book-entry security clearers, conversely, should be little affected by risk reduction procedures. Such securities related daylight overdrafts are—under Federal Reserve policy—to be collateralized, using the underlying security being transferred or a pool of collateralizable securities. No sender net debit caps or bilateral credit limits would apply to such book-entry security transactions. The logic behind treating securities transfers differently from funds transfers is to minimize the potential for disruption in this securities market because of the key role it now plays in the implementation of monetary policy. Consequently, the main argument used against collateralization of funds transfers—that collateralization will not generate strong incentives to reduce the *level* of daylight exposure—is the main argument in its favor when applied to U.S. government securities transfers because the imputed benefits of having a stable government securities market for the implementation of monetary policy is thought to outweigh the benefits of having more stringent risk reduction measures applied to this market.

---

[73] Also, some receivers may not set bilateral net credit limits exclusively on the basis of their assessments of a sender's capital level or risk. Instead, they may set the limits low as a means to inhibit their competitor's ability to provide payments services. Such receivers would be using credit limits as both a risk control and an anti-competitive mechanism.

One other competitive effect deserves mention. Namely, if some sort of payments finality can be achieved on private networks, then their long-term viability *vis a vis* FedWire (which already has finality but has not priced it) is assured. One worry about the present CHIPS system is that if there ever were a significant settlement failure which created losses for users and/or significantly disrupted financial markets, then users would have a strong inducement to shift to networks which have finality of payments. This could possibly result in a permanent loss in CHIPS volume, substantially increasing the costs to remaining users because certain expenditures would be essentially fixed in the short-run.[74]

Pricing Interbank Risk and Cost Recovery. If some form of receiver guarantees are adopted, banks will have the difficult task of pricing this increased risk and devising procedures to pass this cost and the increased internal monitoring/control expenses on to others. These costs may be passed forward to their corporate customers or backward to sending institutions, although the former is often much easier to do in practice.

The increased costs of payments finality and risk reduction on large-dollar networks will lead, in some cases, to the development and use of substitute procedures to minimize these new costs and risks. One likely outcome noted earlier would be the substitution of ACH type transactions for large-dollar payments, once same-day ACH settlement is possible. However, if enough large dollar payments shift to ACH transfers, similar payments finality and risk reduction procedures will probably also be required for ACH transfers as well. Only if small dollar funds transfers over FedWire and/or CHIPS shift to ACH transfers will there likely remain a significant difference in the treatment of daylight exposures on wire transfer *vs.* ACH networks.

The potential for only shifting small dollar funds transfers—those under $100 thousand per transfer—to an ACH network is quite high on

---

[74] CHIPS operating expenses are distributed to users in proportion to volumes sent, a procedure which is equivalent to average cost pricing with no volume discount.

FedWire since some 60 percent of FedWire funds transfer volume falls within this range. The potential for this occurring on CHIPS is low, however, since only around 10 percent of CHIPS volume represents domestic funds transfers; the remaining 90 percent represents international transfers for which no close substitute currently exists (other than use of the SWIFT message transfer network which settles using correspondent balances which may or may not be fully funded prior to settlement).

Perhaps a more long-term development would be the possible substituton of direct corporate to corporate payments for current arrangements whereby corporate payments pass through the banking system. The ACH Corporate Trade Payments progam could serve as an effective model for this development if nondepository institutions had direct access to an ACH network and could hold reserve accounts in the future. Alternatively, payments information and settlement could in principle be handled completely outside of the banking system or through the use of nonbank banks wholly owned by nondepository institutions (if not prohibited by law in the future). The risk issue, in this instance, would move from interbank to intercorporate payments.

Instititutional Changes in Financial Markets. Another possible response to risk reduction efforts concerns institutional change in certain financial markets. Specifically, if bilateral net credit limits are imposed, or if sender net debit caps restrict payments at some banks, or if fees are imposed on the extension of interbank daylight credit, then new institutional arrangements may develop in the federal funds market. Instead of traditionally returning overnight federal funds in the mid-morning and receiving new overnight funds by late afternoon, the following changes could occur:

(1) Increased use of term and continuing contract federal funds to reduce daylight overdraft positions caused by use of overnight federal funds on FedWire;

(2) Return of *net* overnight federal funds when overnight funds are repurchased from the same lender; and

(3) Possible development of an intra-day funds market to reduce daylight overdrafts—but still make payments—on any large-dollar network. This development could lead to increased transaction volume on all large-dollar networks.

These institutional changes, particularly the increased use of term and continuing contract federal funds (now around 25 percent of the total federal funds market), can lengthen liability maturity structures. Such a development, if widespread, could decrease the flexibility and usefulness of liability management as a tool to fund unexpected deposit withdrawals or loan takedowns. Alternatively, if an intra-day funds market develops, this could lead to increased money desk trading opportunities with consequent increases in the interest rate risk of this operation as well as the possiblity for increased profitablity.

One unanswered question concerns the treatment of definitive securities (such as CDs, bankers acceptances, and commercial paper) if book-entry procedures are developed for these types of funds transfer transactions in the future. Should book-entry definitive security transfer caused daylight exposures be treated the same way as book-entry government security overdrafts; namely, permit the use of the underlying security as collateral for the overdraft? If so, this implies a significant change in the current customer-bank legal arrangements regarding bank purchases of these instruments for their customers. In effect, current legal agreements would need to be altered to allow a bank to pledge these securities as collateral if and when an interbank daylight exposure arises when these instruments are traded.

Collateralization of definitive security transfers and foreign exchange contracts could be implemented on CHIPS, reducing the risk of loss (but not necessarily the liquidity risk) in the event of sender failure. Whether such an arrangement would be implemented on Fed-Wire is unknown. One would have to make a convincing argument that, like government security transfers, definitive security transfers serve some higher purpose justifying a different treatment from other funds transfers.

The three institutional changes noted above could be induced by the implementation of relatively low receiver bilateral net credit limits, by adopting low sender net debit caps, or by imposing fees on the extension of daylight credit. The first two policies would act like a position limit or quota and, in effect, absolutely limit the supply of intraday funds to finance intraday transactions. In response, and to provide flexibility, some banks with excess intraday funds (or the Federal Reserve System) could stand ready to lend funds intraday to others for a fee. In this event, an intraday funds market could evolve which would accommodate demands for intraday funds from banks wishing to send transfers in excess of their sender caps. What would be the price established in such a market? No one knows for sure since no intraday market of any size now exists. The only apparent exception would be the dayloan market between banks and securities broker/dealers where the intraday loan rate has been set institutionally at 100 basis points for a number of years.[75]

While a 100 basis point rate may perhaps be an initial price set for intraday funds, if such a market evolves, it is not likely to be the price in the long-run. This is because cheaper substitute procedures are available to reduce daylight overdrafts rather than maintain their current level and fund them with intraday money. Two substitute alternatives were listed above: increased use of continuing contract federal funds and return of net overnight funds when funds are repurchased from the same lender. For the first alternative, the spread between the overnight and the continuing contract rate (some 20 basis points at this writing) would likely be close to the effective "price" of reducing daylight overdrafts resulting from overnight funding operations. The price of the second alternative is unknown but, in informal discussions with bankers, appears to be very much below the dayloan rate of 100 basis points. In sum, the institutional changes likely to be associated with risk reduction on large dollar funds transfer networks are difficult to accurately predict but many choices appear to exist. And none, except delaying sending funds until the end of the day—a solution which can not work for all senders over time—is free.

---

[75] Some exceptions exist since the most risky borrowers in this market have been charged 200 basis points in the past while other very low risk situations have been assessed a rate lower than 100 basis points.

X. Summary and Conclusions

This study attempts to sort out and analyse, in economic terms, some of the effects that the Depository Institutions Deregulation and Monetary Control Act of 1980 (MCA), and related payments developments, have had or may have on the U.S. payments system. The main topics covered include:

Costs:

(1) Estimates of the user and social costs (i.e., all user costs less transfer payments) of cash, checks, ACH (automated clearing house), and wire transfer of funds payment methods;

(2) Estimates of the costs (to banks) and the revenues (to the U.S. Treasury) of the MCA;

Pricing:

(3) Analysis of methods used to price Federal Reserve payment services and float;

(4) Analysis of the so-called Private Sector Adjustment Factor (PSAF) or "markup" to Federal Reserve payment service prices;

(5) Estimates of scale economies in Federal Reserve (wholesale) processing of cash, check, ACH, wire transfer, and book-entry U.S. government security payment operations;

(6) Estimates of scale economies in bank (retail) payment related operations in collecting and purchasing deposits and in making and disbursing loans;

Competition:

(7) Estimates of Federal Reserve and private sector market shares in processing check, ACH, and wire transfer payment instruments;

(8) Analysis of nonprice aspect of payments services, and potential new payment arrangements, which can affect competitive structure in the future;

Risk:

(9) Analysis of current developments regarding large dollar wire transfer networks as a result of efforts to reduce risk of a settlement failure; and

(10) Analysis of the likely market response to final adoption of bilateral net credit limits, sender net debt caps, payments finality, or assessment of fees for the provision of daylight credit on large dollar payment networks

All of these topics, save for the estimation of Federal Reserve and commercial bank scale economies (topics 5 and 6, above), have not been heretofore publicly addressed or only addressed in a very limited way.[76] Generalizing from the analyses presented and discussed above, the following conclusions regarding costs, pricing, competition and risk in the U.S. payments system are offered.[77]

The Costs of Different Payments Instruments. Most industry observers are well aware that the loss of check float benefits to check writers is one of the main reasons why check users (both business and consumer) have not, on the whole, shifted to no float (or low float) payment alternatives. No float (or low float) alternatives are cash, ACH, debit card, POS, or wire transfers. What is not so well-known is that the float benefits to check users, in the aggregate, is some $33 billion a year (Dudley, 1983) and, according to our analysis, *exceeds* the real costs of producing, handling, processing, and transporting checks to complete a payments cycle. This result, for the nation as a whole, closely mirrors a more detailed and independent analysis of the U.S. government's real resource and float costs of using a check (Dudley, 1984). Thus our numerical analysis, which can no doubt be improved upon, is robust. That is, two independent analyses reach the same conclusion: the benefits of check float are larger than all the other costs in using checks. This result means that the net cost of using a check for the check writer is negative. From the user's point of view, which includes the costs (for cash) and benefits (for check and ACH) of float, the unit user cost per transaction with cash, checks, and ACH are respectively, 19¢, −15¢, and 33¢. The use of a check as a means of payment, of course, reflects the influence of these relative prices.

But is the current level of check use in society's best interest? That is, what are the unit costs per transaction for society as a whole where the float transfer payments net themselves out? For society as a whole, the float benefits to check and ACH users (payors) are offset by the the float costs of those who accept checks for payments (payees). As well, the "float" or opportunity costs to holders of non-interest earning cash balances are offset by the seigniorage or money producing benefits to the government. Because the transfer payments offset one another for society as a whole, the unit social costs per transaction for cash, checks, and ACH become 7¢, 68¢, and 34¢, respectively. From this perspective, that is looking only at the real resources involved (and excluding transfer payments), checks are twice as expensive as the next best payment alternative. In economic terms, real resources are being misallocated because the proper price incentives (the unit social costs) are quite different from the price incentives faced by those choosing to use one payments method over another (as the decision maker faces unit user costs, not unit social costs). This situation is a case of "market failure", since the price system does not now allocate resources properly and everyone can (in principle) be made better off by changing the current price signals to more closely reflect their true social cost.

---

[76] The most comprehensive public discussion and presentation of information to date is contained in a publication which resulted from Congressional hearings on *The Role of the Federal Reserve in Check Clearing and the Nation's Payment System* (U.S. Congress, 1983).

[77] These analyses and conclusions should not in any way be attributed to the Board of Governors of the Federal Reserve System or its staff. They represent a purely personal view of the author alone.

Two things would have to change for this to occur. One is clearly possible while the other is very unlikely. First, the Uniform Commercial Code would have to be altered and/or federal legislation passed allowing payees to shift their check float costs to payors, who create the float to begin with. Such an arrangement could be implemented by allowing check processors—correspondent banks, service bureaus, and the Federal Reserve—to charge payor banks for the float involved in collecting checks deposited for collection by payees. The revenues from these payor float charges, in turn, would be passed back to the collecting bank and subsequently, to payees. While this may not entirely eliminate the benefits of check float to payors, as mail float may still exist, it would substantially reduce these benefits and result in a relative price structure which more closely reflected the cost to society of using different payment instruments. In particular, shifting the cost of float to those payors who create it would significantly reduce the wasted real resources involved in remote disbursement and special expedited collection in order to maximize check disbursement float or minimize receivables float.

The second, and very unlikely, change required to better allocate payment resources would be for the government to give up the approximately $15 billion a year in seniorage benefits obtained by being the sole provider of lawful coin and currency and offset these lost revenues elsewhere through the tax structure. This, of course, is the economic perscription, one that for political reasons (and, perhaps, a preference for user fees as opposed to income or sales taxes) is most unlikely to occur in practice.

It is possible to illustrate the savings in real resources which could be associated with charging payors for the check float they create. For example, if payors lost all their float benefit, they would face the social costs of using a check (68¢) and would be induced to shift, over an appropriate adjustment period, to the ACH which costs only one-half as much (34¢). If all 40 billion check transactions shifted to ACH, the savings in real resources, assuming nothing else changes, may be on the order of $13.6 billion a year (i.e., (68¢ - 34¢) x 40 billion = $13.6 billion). Of course, not all check users would shift to an ACH no matter what the cost difference. Also, ACH and check costs may vary in ways that are not now foreseeable, although at present checks face constant costs and ACH has economies of scale at current volumes (which are now only at 1 percent of total check volumes). The point remains, however, that continued use of checks is wasteful and, at least on economic grounds, a strong case can be made to permit payees to shift the float costs created by payors to the payors. Such an arrangement is both equitable and efficient. It is also feasible if phased-in over time.

The Effect on Bank Costs from the Monetary Control Act (MCA). Although not well-known, the net effect of the MCA has been to raise Treasury revenues compared to what they have been estimated to be without the Act. This conclusion was presented to Congress prior to the passage of the Act in 1980, was published in the *Congressional Record*, and has been verified by more recent and complete estimates provided to the Treasury every year since then. The MCA changed four things that affected Treasury revenues:

(1) It lowered reserve requirements on Federal Reserve member banks;

(2) It imposed reserve requirements on all nonmember banks and other depository institutions holding transactions balances;

(3) It required that the Federal Reserve price its payments services; and

(4) It required that Federal Reserve generated float be priced and/or eliminated.

The net (after tax) effect of these changes is estimated to have raised Treasury revenues by an average of $550 million a year over the eight year period 1981-1988 that reserve requirements are being phased-in for non-member depository institutions. It is important to note that most of this estimated gain—some 75 percent of it—arises from preventing a fall in Treasury revenues from the expected attrition in Federal Reserve membership (and therefore lost revenues from reductions in idle reserve balances) had not the MCA been passed. Thus, compared to the preexisting level of Treasury revenues from idle reserve balances and the cost of providing "free" payments services to member banks prior to the MCA, the realized increase in Treasury revenues is only 25 percent of $550 million a year; the other 75 percent comes from not realizing a reduction in Treasury revenues which may otherwise have occurred.

Scale Economies and the Pricing of Federal Reserve Payment Services. The Federal Reserve prices its payment services in much the same way that large correspondent banks price theirs. That is, both cost and the presumed price sensitivity of the user are taken into account in determining individual service prices (e.g., mixed deposit, group sort, fine sort, etc.) within a service line (e.g. checks). The Federal Reserve, however, prices these individual services so that the revenues raised are expected to equal the total costs involved for the entire service line (including the so-called private sector adjustment factor or PSAF). This is done separately for each service line (e.g., cash, checks, ACH, wire transfers, etc.). There is no cross-subsidization between service lines. That is, as a matter of explicit policy, excess revenues are not generated in, say, wire transfers, to offset a revenue shortfall in, say, ACH or checks. Such a narrow approach to pricing is, we understand, infrequently followed at large correspondent banks who offer payment services in competition with the Federal Reserve. In constrast, correspondent banks typically price to make the entire package of services offered to respondents profitable. Losses or lower than desired profits on one service line may be made up elsewhere in the package of services supplied and service line cross-subsidization thus becomes a common occurrence. Such an arrangement, however, is economically justified either across service lines or within them.

A pricing method which considers both costs and demand elasticity and matches revenues with costs, called Ramsey pricing with a revenue constraint, has been shown to be an optimal departure from marginal cost pricing. If no revenue constraint existed, marginal cost pricing would be the preferred pricing method for economic efficiency reasons. The MCA instructed the Federal Reserve to match revenues with costs (plus the PSAF) and thus specified that a revenue constraint be met. There are two ways to do this:

(1) Average cost pricing; or
(2) Ramsey pricing with a revenue constraint.

The Federal Reserve used average cost pricing for one year then switched to Ramsey pricing once enough experience was gained to make this possible. Applications of market sensitive pricing to checks and wire transfers and benefit-flow pricing to ACH represent Federal Reserve terminology for Ramsey pricing with a revenue constraint. When scale economies exist or if constant costs occur, the Federal Reserve's market sensitive pricing approach is well-justified economically. Real resources will be used most efficiently as a result. But when scale diseconomies exist, as apparently occurs for cash services or occurred for checks prior to pricing, the market sensitive pricing approach is not optimal unless it is altered to conform to how Ramsey pricing would be applied. Currently, this issue is moot since very few cash services are priced and scale diseconomies for checks have, after pricing, been eliminated. Instead, constant costs now exist for checks so market sensitive pricing is appropriate for all Federal Reserve payment services. In sum, knowledge of scale economies (or how average costs vary as cash, check, ACH, and wire transfer processed volumes are increased) is essential in choosing and defending a pricing policy. An optimal pricing policy for the Federal Reserve should do three things:

- Price so that revenues match costs (plus the PSAF)—the revenue constraint;
- Price to maximize volume produced when scale economies exist (i.e., when average costs fall as output rises); and
- Price to minimize volume produced when scale diseconomies occur (i.e., when average costs rise as output rises).

When constant costs exist, market sensitive pricing, marginal cost pricing, and average cost pricing are all equally efficient, although they may have differing effects on the amounts actually produced.

The essense of market sensitive pricing is to charge prices somewhat higher than average costs for those individual services within a service line that face relatively inelastic demands and charge prices somewhat below average costs for those individual services facing relatively elastic demands. At no time are prices to be set below average variable costs (an approximation for marginal costs, which are difficult to measure). Since these prices are adjusted to ensure that revenues match costs (plus the PSAF), the MCA required revenue constraint is met. And when scale economies exist, as can be demonstrated for ACH and book-entry security transfers, this pricing procedure is an optimal departure from marginal cost pricing. When constant costs exist, as for wire transfers and checks (after pricing), market sensitive pricing yields the same economic efficiency results as marginal cost pricing.

While statistical studies have been completed which estimate Federal Reserve scale economies in most of the various payment services provided, no comparable studies have been done, or at least none made publicly available, for payment services supplied by correpondent banks. Internal analyses at one bank over time do not qualify. In order to reach statistically valid results which can support general conclusions, it is necessary to have information on the cost-output relationship for more than one or just a few banks and, in addition, hold constant other influences on costs which would otherwise bias these results (e.g., control for changes in labor costs, different check sort patterns, etc.). This has not been possible for two reasons:

(1) Many commercial banks do not collect cost data in a form appropriate for statistical modeling; and

(2) Those banks which do have the appropriate cost accounting data have viewed it as proprietary in nature and have not made it publicly available.

The closest one can now come to determining the scale economies which may exist in commercial bank payment operations is to measure the scale economies which are associated with bank deposit taking and loan making activities. There have been many studies of scale economies in banking and the conventional wisdom is that scale economies exist for the bank as a whole, as well as for many of the individual activities. More recent studies have challenged this view, however. It has been demonstrated—with larger samples of banks, more recent data, improved statistical modeling, and correcting for an important measurement error—that if anything scale diseconomies or constant costs are the rule at commercial banks. The only consistent exception to this result occurs at small banks where scale economies exist for those institutions with less than \$50-\$75 million in deposits. These results are illustrated in the text.

Clarifying the Private Sector Adjustment Factor (PSAF). The PSAF was contained in the MCA. Its purpose is to promote competition between the Federal Reserve and correspondent banks by, in effect, requiring that Federal Reserve payment service prices contain a "profit" element in addition to the basic costs of production. There are three essential elements to the PSAF:

(1) Tax rates;
(2) Capital structure (the debit-equity ratio); and
(3) Return on capital (the weighted average of debt costs and return on equity).

The basis for these three components has been, and continues to be, a set of large commercial banks. Many adjustments are made, however, before these data are used in the PSAF. Bank tax rates are adjusted downward to remove taxes incurred but not actually paid and adjusted upwards to account for nontaxable income from bank holdings of state and municipal securities. Large banks have recently faced a (marginal Federal) tax liability of around 42 percent but, of course, end up paying a considerably lower percentage for a variety of reasons. The (average) tax rate proposed in the PSAF for 1984 is 36 percent.

In terms of capital structure, large banks are highly leveraged with equity capital comprising only 5 or 6 percent of total assets. By "matching" the life or type of priced service assets with different financing methods, the Federal Reserve has adopted a capital structure composed of short-term debt (9 percent), long-term debt (27 percent), and equity capital (64 percent). The corresponding rates of return proposed were 9.5 percent, 10 percent, and 21 percent (pre-tax), respectively. These rates of return were derived from data in annual reports of large bank holding companies.

The weighted average cost of capital implied here is 17 percent. This figure is applied to the priced service assets requiring financing and generates a PSAF value of $56 million. This $56 million is added to the $79 million in transportation expenses and the $389 million in Federal Reserve direct, support, and overhead costs, making a total of $524 million to be recovered in pricing for all priced payments services. This process has, unfortunately, led to a considerable amount of public confusion. First, since the $56 million in financing costs can be viewed as a "mark-up" of some 14.5 percent on the direct, support, and overhead costs of $389 million, some have mistakenly confused this 14.5 percent figure with the average cost of capital (which is in fact 17 percent) or even the pre-tax return on equity capital (which is 21 percent). (The reasons for excluding transportation costs from the expenses to be marked-up are discussed in the text.) A second confusion has arisen when others mistakenly compare this 14.5 percent "mark-up" with mark-up pricing practices in the private sector. Such a comparison is not appropriate because the latter includes overhead expenses and financing costs in the markup percentage while the Federal Reserve "mark-up" only includes financing costs. If the traditional mark-up approach were applied to Federal Reserve payments services, the mark-up percentage would be 69 percent (since both the $56 million in financing costs plus the $125 million in overhead expenses would represent the mark-up over the $264 million in direct and support costs, mirroring private sector practice).

Overall, much effort has been expended in trying to fairly interpret and implement the MCA requirement to include a PSAF in the prices for Federal Reserve payment services. We have attempted to clarify some of the more common misconceptions regarding the PSAF's calculation. Of a more fundamental nature are criticisms—not misconceptions—concerning the use of a set of large banks in determining (after appropriate adjustment) the three elements which make up the PSAF calculation (tax rates, capital structure, and cost of capital). Certain groups of correspondent banks, the primary competitors of Reserve Banks in supplying payment services, have argued that nonbank data processing companies provide a more appropriate basis upon which to build a PSAF. The basic issue is thus which group, a set of a large banks or a set of nonbank data processors, provide a closer match to those payment services provided by the Federal Reserve. Large banks provide correspondent bank payment services to their respondent bank customers as only *one part* of their total service offering while nonbank data processors only provide *some types* of correspondent bank payment processing and accounting services for banks. The first group, in effect, provides "too many" services while the second provides "too few' compared to those offered by the Federal Reserve. Thus, regardless of the model chosen for the PSAF, substantial adjustments are necessary in applying it to the Federal Reserve. Because there are problems with using either large banks or nonbank data processors as a model for the PSAF, this issue cannot be easily resolved and remains a matter of judgement. This judgement will likely be determined for most by assessing the reasonableness of the adjustments made using the large bank model for the PSAF. These adjustments were outlined above and are discussed in more detail in the text.

Market Shares and Competition in Providing Payment Services. The Federal Reserve's estimated share of the market for (wholesale) check, ACH, and wire transfer services are 33, 95, and 67 percent, respectively. Prior to pricing, these estimated shares were 43, 95, and 64 percent. Thus pricing has reduced the Federal Reserve's share of the check processing and collection market by 25 percent, reducing it from 43 to 33 percent. There has been little or no change in the ACH or wire transfer market shares due to pricing.

The ACH market share remains high because there are few private sector suppliers, other than the New York Clearing House Association. Perhaps one reason for this result is that pricing for the ACH service, unlike that for checks and wire transfers, was not at full cost when pricing began in 1981. ACH pricing is being phased-in over a five year period so that only in 1986 will ACH prices fully reflect all the costs involved. As ACH prices reflect more costs, it is possible that the Federal Reserve's market share will fall as new suppliers enter this market.

The scale economies found to be associated with ACH processing would ordinarily be thought to limit new entry by private sector suppliers. However, because of the unconsolidated nature of the Federal Reserve's ACH processing operations (with some 35 separate processing "centers"), this may not prove to be the barrier to entry it is normally found to be. New entrants, if they are much more consolidated than the existing Federal Reserve structure, may be able to realize a greater portion of ACH scale economies and so face lower costs per transaction over time.

In wire transfers, a different situation exists. The likely long-run impact of pricing, along with other changes, may be to shift wire volume to the ACH (once same-day ACH settlement is offered, probably in 1985). Small-dollar wire transfers, those for less than

$100,000, account for around 58 percent of the Federal Reserve's wire transfer volume and these transfers would be the most likely candidates for such a shift. The price differential is very attractive here—65¢ to initiate a wire transfer versus less than 5¢ (today) to initiate an ACH transfer.

Development of new payment services can also influence market share, both between different suppliers and between different payment methods. Check truncation, where the physical check is kept at the bank of first deposit or at some other point in the collection process, should foster electronic payments. This should expand the ACH or some new electronic system similar to it. Electronic check collection (ECC), if targeted as expected for large-dollar checks, should foster wire transfers. ECC would reduce or eliminate the float benefits of using checks for very large-dollar payments (those over, say, $100,000). Because of the large values involved, these check payments would more likely shift to wire networks, which are more secure, than to the ACH. Lastly, if a direct settlement service (DSS) is offered, there could be a shift in check volume from one processor to another, as well as a reduction in remote disbursement. DSS would permit correspondent banks to present checks directly to payor banks (using their own transportation), require payment in immediately available funds, and not be subject to presentment fees. The usefulness of ECC and DSS have been assessed in terms of a social cost/benefit analysis. Such an analysis, correctly, looks at the net effect on real resources from the point of view of society as a whole and thus any float benefits or costs net each other out. On the basis of these and other considerations, ECC was not adopted by the Federal Reserve. DSS, which was proposed to the Federal Reserve by correspondent banks, is still under study.

Perhaps the most important influence on market shares and competition, however, will be interstate banking. By increasing the size and geographical dispersion of banking organizations, interstate banking will increase banking concentration and result in larger volumes of "on us" check, ACH, and wire transfer payments. "On us" payments only require the internal movement of funds between accounts within the same banking organization, rather than the external movement of funds between accounts at different banks. Greater banking concentration will also lead to a consolidation of check presentment endpoints making direct presentments by banks more cost-effective. In turn, these developments should increase incentives for direct check exchanges between local/regional clearinghouses and between originators and receivers on private sector ACH and wire transfer networks.

The end result to interstate banking is seen in Canada, England, and Europe where often the top 10 banks control over 90 percent of all deposits. In these countries the central bank provides no payments processing services (except for net settlement). All payments transactions are handled internally at banks or at very large national payments clearinghouses, or by the Postal Service (which runs the European giro system). While the U.S. banking system is not likely to be so concentrated, the concentrations which can be expected would significantly reduce the role of the Federal Reserve, large correspondent banks, and nonbank service bureaus who provide payments processing and bank accounting services. The externally handled processing volumes of today will shift to internal transfers at the largest banks and new direct payment exchange networks will emerge.

Current Developments in Reducing the Risk of a Settlement Failure. Risks exist in any payments system. There are risks due to fraud, accounting errors, operational breakdowns, and the unexpected failure of a payments network participant to settle for funds transferred during the day. The risk issue discussed here is settlement risk, along with the various methods which have been suggested to control and reduce this risk.

Settlement risk exists for all types of payments networks but its impact on financial markets would be greatest if it occured on a large-dollar wire transfer network. These networks transfer up to $500 billion a day and have an average transaction value of $2.5 million per transfer. In the current operation of wire transfer networks, both those of the Federal Reserve (FedWire) and the private sector (CHIPS, CashWire, CHESS), daylight overdrafts are incurred by certain large and some small banks. Should a wire transfer network participant (unexpectedly) fail to settle for funds transferred over these networks while in a daylight overdraft position on a network, this single failure to settle can lead to the settlement failure of other participants on the same network in a systemic fashion. Thus one bank's daylight overdraft—a credit risk to those participants receiving funds from this bank—can create systemic risk for all other participants on the network. This is a result of the interconnectedness which exists, directly and indirectly, between participants on a wire transfer network. By reducing or controlling daylight overdrafts, systemic risk is reduced, thus reducing the likelihood of financial market disruption from a failure to settle.

Daylight overdrafts basically represent an extension of daylight credit to cover the net debts experienced by certain banks which send a larger value of wire payments than they receive, at least for a few hours during the day. Daylight overdrafts are typically erased by the end of the day when the flow of incoming funds rises and eliminates the overdraft and/or settlement occurs which removes the net debit built up over the day.

The proximate cause of funds transfer daylight overdrafts on the major wire transfer networks (FedWire and CHIPS) concerns the current institutional practices which have evolved in the overnight funds markets. Daylight credit has been free and, as a result, has been "overused" compared to its true value to the user and the risk to the supplier (and/or society as a whole). This is another case of "market failure" where the price charged for daylight credit (zero) does not properly reflect its (risk or real resource) cost. Change these incentives, by pricing the daylight credit provided or otherwise limiting the value of credit which can be incurred, and participants will alter their current behavior and reduce daylight overdrafts or those providing the credit will be compensated for the risk they incur.

Some of the most often mentioned "solutions" to the daylight overdraft problem involve quantitative limitations on the size of these overdrafts (bilateral net credit limits, sender net debt "caps") or proposals to provide daylight credit at a price. Certain quantitative limitations are already in place, such as bilateral net credit limits on CashWire, CHESS, and CHIPS (late 1984) or sender net debit caps on CashWire, CHESS, and FedWire (as counseling guidelines, not absolute limits). Finality of payments, which puts "teeth" into bilateral net credit limits, only exists on FedWire and CashWire at present. All of these solutions, as well as those noted in the text, are under study by the Federal Reserve and groups of concerned industry representatives. There are many problems with most or all of the solutions put foreward to date but a consensus is likely to emerge in the near future. The most likely outcome, in our view, would be:

(1) Sender net debit caps, applied on each network separately or applied to all networks together, which would place an upper limit on the daylight overdraft *level* a bank could incur at a zero price;

(2) Bilateral net credit limits set by receiving banks on sending banks (using credit officers) which affects the *distribution* of a sender's daylight overdraft across different receivers;

(3) Some form of payments finality or receiver guarantee to a receiving bank's customers that funds received can be used without recourse to the customer in the event the sending bank fails to settle; and

(4) Some type of "overline" credit may be provided at a fee by banks or the Federal Reserve to provide flexibility for the quantitative limitations on funds flows noted in (1) and (2).

Market Response to Risk Reduction on Large-Dollar Networks. Adoption of one or more of the above noted four settlement risk reduction alternatives—sender net debit caps, receiver set bilateral net credit limits, partial or full payments finality, and pricing daylight credit—can significantly affect banks and financial markets in the future. Such possible effects concern:

(1) Changes in the structure and growth of large-dollar wire transfer networks;

(2) Shifts in the competitive postion among banks, depending on which payments services they specialize in;

(3) Improvements in bank monitoring of customer daylight balances and interbank funds transfer positions; and

(4) Changes in the institutional structure of financial markets.

Limiting and/or pricing daylight overdrafts can have the effect of reducing the over 10 percent annual volume growth experienced by wire transfer networks and may reduce incentives for new suppliers to enter this market compared to a situation where daylight credit is free and few limitations exist.

Pricing and quantitative limitations can also affect the competitive positions of those banks which have been frequent users of daylight credit and, in the opinion of some, have incurred excessive daylight overdrafts relative to their capital level. Because of the proposed differential treatment of funds transfer daylight overdrafts and U.S. government book-entry securities related daylight overdrafts, which are to be collateralized because of the importance of this market for monetary control, banks which have specialized in securities transactions may be impacted less than others. Those banks which have chosen to specialize in third party funds transfers, and/or interbank trading in overnight funds in excess of their need for these funds for their own internal use in asset funding, will likely be impacted the most. And, of these banks, those in unit banking states may be more significantly constrained than banks in branching states. Unit banks typically rely more heavily on overnight funds than do branching banks, which use core deposits developed internally within their branching network and often incur lower net positions in overnight purchased funds. These differential impacts by type of payments specialization, by involvement in intraday trading for only trading (not funding) purposes, and by unit or branch organizational structure, will all lead to different costs being experienced and therefore affect the competitive structure between banks. The costs associated with risk reduction, of course, will over time be directly passed on to the various internal bank profit centers affected and, subsequently, to banking customers. This shifting process is made easier if a price for daylight overdrafts develops than if only quantitative limitations are adopted (voluntarily by banks or, perhaps, imposed by bank regulators). Such a situation would be similar to that which occurred after Federal Reserve payments services were priced. Pricing led correspondent banks to reexamine and reprice most or all of their payment services, at times changing their price by more than that attributable to Federal Reserve prices alone.

Lastly, risk reduction could lead to certain changes in financial markets. These changes would likely concern the development of new procedures and increased reliance on existing methods which reduce the impact of overnight funding on the measurement of daylight overdrafts. Banks that typically purchase overnight funds from the same respondent institutions can attempt to bring these funds "in house" using book-entry procedures or a continuing contract. Alternatively, borrowers could return the *net* value of borrowed overnight funds rather than sending an a.m. return for the gross amount borrowed the previous day with a receipt in the p.m. of the gross amount reborrowed from the same lender. Still other changes may involve a shift from overnight funding to term funds or the payment of basis points by borrowers to induce lenders to send borrowed funds earlier than now typically occurs (say one hour after the amount and rate for that day have been negotiated by phone). Alternatively, borrowers could compensate lenders for delayed return of funds borrowed the previous day. This would provide for a closer matching in time for gross sends and receives and reduce the duration of daylight overdrafts and, in some cases, their amount as well. The basic result here is to provide for a better match between a bank's book position regarding funds purchased and sold with the actual operational position regarding funds received and sent.

A more problematical development would be the emergence of an interbank intra-day funds market to cover overdraft positions. In effect, this development is very similar to borrower payment of basis points to lenders to speed up the sending of newly borrowed funds or for the borrower to delay return of funds previously borrowed. Here the supplier of what would otherwise be intra-day funds is the lender himself, rather than some other market participant. The development of a formal and large intra-day funds market, it is noted, faces significant operational difficulties, not in negotiating the transfer of intra-day funds but in actually operating to transfer these funds in a timely manner.

Appendix

A Translog Cost Function Model For Estimating Federal Reserve Payments Processing Scale Economies

The Translog Specification. The translog cost function is a quadratic function and in this application can be expressed as:

$$(1) \qquad \ln TC = \alpha_{TC} + \alpha_Q \ln Q + \alpha_{QQ} 1/2(\ln Q)^2 + F(w_i)$$

where:

TC = total check, ACH, or wire transfer activity or production costs (which excludes certain overhead costs);[78]

Q = total check, ACH, or wire transfer processed volume (which excludes fine sort volume);[79] and

$F(w_i)$ = other influences on costs in (1) in terms of their level, their level squared (a quadratic term), or as an interaction with Q or themselves.

The function $F(w_i)$ differs between the estimated check, ACH, and wire transfer cost functions due to availability of data and the specific production process being specified:

For Checks:[80]

$$F(w_i) = \beta_{PL} \ln P_L + \beta_{IP} \ln IP + \beta_{PL,IP} \ln P_L \ln IP + \beta_{PL,Q} \ln P_L \ln Q + \beta_{PL,PL} 1/2(\ln P_L)^2$$

---

[78] Total activity costs are those fixed and variable expenses that can be closely associated with processing operations. These include labor, equipment, data processing, building depreciation, and shipping costs, plus 11 other accounting categories. Certain other overhead expenses are not included in total activity costs because their allocation to check and other priced services would have to be based on more approximative methods, such as value or expense ratios. Examples of such excluded overhead expenses would be food and canteen services, personnel department costs, and protection.

[79] ACH or wire transfer fine sort volume does not now exist, although it may in the future for ACH. The check volumes used in this study exclude fine sort volume for two reasons. First, fine sort volume is not processing volume; processing is done in the private sector with Reserve Banks only providing transportation, adjustments, etc. Second, the fine sort activity comprises less than 4 percent of the check total activity costs and is therefore small. Reserve Bank check volumes, included checks handled by two Reserve Bank offices, around 1.5 billion items out of the 13 billion items handled overall.

[80] Certain additional control variables, such as holdover float value and the number of internal errors per 1,000 checks processed, could have also been used in 1982 and 1983, but not in 1976 when an earlier and more sophisticated check study was done. Since one of our purposes here is to contrast scale economies both pre and post FR pricing, variables which were not available in the pre-pricing period were not used in the post-pricing specification.

(2) *For ACH:*[81]

$$F(w_i) = \gamma_{PL} \ln P_L + \gamma_G \ln G + \gamma_{IE} \ln IE + \gamma_E \ln E$$

For Wire Transfers:[82]

$$F(w_i) = \delta_{PL} \ln P_L + \delta_{LT} \ln LT + \delta_{ON} \ln ON$$

where:

PL = average (unit) labor costs incurred in check, ACH, or wire transfer operations;

IP = the check item pass ratio, which helps to capture differences in check-sort patterns across offices;

G = proportion of total ACH volume which is government originated;

IE = number of internally generated processing errors per 1,000 ACH images originated, a quality measure;

E = percent of endpoints receiving (more expensive) magnetic tape deliveries rather than electronic delivery;

LT = percent of wire transfers that were late, a quality measure; and

ON = percent of transfers made with on-line equipment, which are less expensive to process than using the (off-line) telephone.

The prices of capital structure and equipment are not included in these specifications for three reasons. First, developing these data properly is very, very difficult. Second, capital structure costs are a very small component of total production costs—around 5 percent for checks and 2 percent for ACH and wire transfers. And while the capital equipment costs are larger—approximately 15 percent for checks and wire transfers, and 30 percent for ACH—their prices would show little or no variation in the cross-section data set since standard, nationally available computers and sorting machines are being used across offices. And third, when a price of capital was constructed and used in an earlier check study, the estimated parameter used in the scale economy calculation was very small (so that the scale economy estimate is very unlikely to be importantly affected one way or the other). For these reasons, the model relies on only one factor input (labor) and neglects the other (capital).

---

[81] The full ACH translog model incorporated 10 additional variables composed of squared and interaction relationships among the five ACH variables shown in (1) and (2). Due to high collinearity among these variables, these was no statistically significant difference in the model's overall explanatory power when these 10 additional variables were excluded. As explained below, the ACH model shown here can be simplified even further and not significantly affect the model's explanatory power.

[82] The full wire transfer model also originally included more variables but it too, like ACH, could be simplified without significant loss of explanatory power.

Measurement of Scale Economies. A measure of scale economies derived from (1) would be:

$$\text{SCE} = \partial \ln TC/\partial \ln Q = \alpha_Q + \alpha_{QQ} \ln Q + \partial F(w_i)/\partial \ln Q. \tag{3}$$

In all the results obtained using (1) it turns out that the cost model can be simplified by neglecting $F(w_i)$ for all three estimations without having a statistically significant effect on the explanatory power of the model. Indeed, in the ACH estimates, both $F(w_i)$ and the quadratic output term $(1/2(\ln Q)^2)$ can be neglected.[83] For our purposes, the scale economy calculation can be simplified to:

For Checks:

$$SCE_C = \alpha_Q + \alpha_{QQ} \ln Q$$

(4) *For ACH*:

$$SCE_A = \alpha_Q$$

For Wire Transfers:

$$SCE_W = \alpha_Q + \alpha_{QQ} \ln Q$$

The SCEs computed are based on statistical estimates of (1) after it has been simplified to $\ln TC = \alpha_{TC} + \alpha_Q \ln Q + \alpha_{QQ} 1/2(\ln Q)^2$ for checks and wire transfers and $\ln TC = \alpha_{TC} + \alpha_Q \ln Q$ for ACH. Estimation of these simplified cost functions did not significantly reduce the model's explanatory power (at the 99 percent level of confidence) compared to estimation of the more complex model specified in (1).

---

[83] The reasons why these extra terms, as well as others not shown here, can be neglected is that: (1) the terms in F(wi) are only weakly related to $\ln TC$ in the check, ACH, and wire transfer data; and (2), the quadratic term $½(\ln Q)^2$ in the ACH data is highly collinear with $\ln Q$ (the $R^2 = .99$) so that no significant explanatory power is lost when this term is neglected. Put differently, $\ln Q$ by itself explains around 75 percent of the variation in $\ln TC$ for the ACH model while $\ln Q$ and $½(\ln Q)^2$ explain around 90 percent of the variation in $\ln TC$ in the check and wire transfer models. Thus the additional variables, when added to the model, add little significant additional information.

Bibliography

American Bankers Association, *Developing Bank Services: A Practical Guide*, Washington, D.C. (1982).

Association of Reserve City Bankers, *Report on the Payments System*, Washington, D.C. (April 1982).

———, *Risks in the Electronic Payments Systems*, Washington, D.C. (October 1983).

———, *Final Report of the Risk Control Task Force*, Payments System Committee, Washington, D.C. (October 1984).

Bank Administration Institute, *Payment Systems in Eleven Developed Countries*. Park Ridge, Illinois (1980).

Baumol, William J., and David F. Bradford, "Optimal Departures from Marginal Cost Pricing," *American Economic Review*, *60* (June 1970), 265-89.

Bell, Frederick W., and Neil B. Murphy, *Costs in Commercial Banking: A Quantitative Analysis of Bank Behavior and Its Relation to Bank Regulation*, Research Report No. 41. Boston: Federal Reserve Bank of Boston (1968).

Belton, Terrence M., "Economies of Scale in Book-Entry Securities," mimeo, Financial Studies Section, Board of Governors of the Federal Reserve System (forthcoming).

Benston, George J., *Federal Reserve Membership: Consequences, Costs, Benefits and Alternatives*, Association of Reserve City Bankers, Chicago, Illinois (September 1978).

———, Gerald Hanweck, and David Humphrey, "Scale Economies in Banking: A Restructuring and Reassessment," *Journal of Money, Credit, and Banking, 14* (November 1982, Part 1), 435-56.

Berlin, E., C. Cicchetti,and W. Gillen, *Prospective on Power: A Study of Regulation and Pricing of Electric Power*, Appendix B, Ballinger, Cambridge, MA, (1974).

Board of Governors of the Federal Reserve System, *Functional Cost Analysis*. Washington, D.C. (1982).

———, *Planning and Control System Expense Report* (PACS), Washington, D.C., detailed annual reports, various years.

Burke, James, "Antitrust Laws, Justice Department Guidelines, and the Limits of Concentration in Local Banking Markets," *Staff Studies*, Board of Governors of the Federal Reserve System, Washington, D.C. (June 1984).

Clark, Jeffrey A., "Estimation of Economies of Scale in Banking Using A Generalized Functional Form," *Journal of Money, Credit, and Banking, 16* (February 1984), 53-68.

Chicago Clearing House Assocation, *Daylight Overdrafts*, Chicago, Illinois (1984).

Cox, William, and Paul Metzker, "Displacing the Check," Federal Reserve Bank of Atlanta *Economic Review, 68* (August 1983), entire issue.

Curtin, Richard T., *Payment Method Costs: Assessments by Retailers*, Survey Research Center, University of Michigan (September 1983).

Dudley, William C., "The Tug-of-War Over 'Float,' " *Morgan Guaranty Survey*, Morgan Guaranty Trust Co., New York (December 1983).

———, "A Comparison of Direct Deposit and Check Payment Costs," *Staff Studies*, Board of Governors of the Federal Reserve System, Washington, D.C. (October 1984).

FAC/TIAC, *Report of the Federal Advisory Council/Thrift Institutions Advisory Council Payments System Committee*, November 2, 1983 (subsequently reprinted in the *American Banker*, November 30, 1983, 10).

Federal Reserve Bank of Atlanta, *A Quantitative Description of the Check Collection System*, Volume 1. Atlanta (1979).

Frodin, Joanna, "Fed Pricing and the Check Collection Business: The Private Sector Response," Federal Reserve Bank of Philadelphia *Business Review*, (January/February 1984), 13-24.

Gambs, Carl M., "The Cost of Operating the Payment System," Proceedings of a Conference on *The Future of the U.S. Payments System*, Federal Reserve Bank of Atlanta, Atlanta, Georgia (June 23-25, 1981), 133-40.

Garbade, Kenneth D., and William L. Silber, "The Payment System and Domestic Exchange Rates: Technological Versus Institutional Change," *Journal of Monetary Economics, 5* (January 1979), 1-22.

Goldberg, Lawrence G. and John T. Rose, "The Effects on Nonmember Banks of the Imposition of Member Bank Reserve Requirements—With and Without Federal Reserve Services," *Journal of Finance, 31* (December 1976), 1457-69.

Guiltinan, Joseph P., *Pricing Bank Services: A Planning Approach*, American Bankers Association, Washington, D.C. (1980).

Humphrey, David B., "Economies of Scale in Federal Reserve Check Processing Operations," *Journal of Econometrics, 15* (January 1981), 155-73.

———, "Costs, Scale Economies, Competition, and Product Mix in the U.S. Payments Mechanism," *Staff Studies*, Board of Governors of the Federal Reserve System, Washington, D.C. (April 1982).

———, "Costs and Scale Economies in Bank Intermediation," in R. Aspinwall and R. Einsenbeis (eds.), *Handbook For Banking Strategy*, Wiley & Sons, New York (1985).

Koot, Ronald S., and David A. Walker, "The Mark-Up in Pricing of Federal Reserve Services," *Issues in Bank Regulation, 4* (Summer 1980), 21-6.

Kopecky, Kenneth J., Darrel W. Parke, and Richard C. Porter, "Framework for Analyzing Money Stock Control Under the Monetary Control Act," *Journal of Economics and Business, 35* (June 1983), 139-57.

Lipsey, Richard G., and Kelvin Lancaster, "The General Theory of Second Best," *Review of Economic Studies, 24* (January 1956), 11-32.

Miller, Preston J., "The Right Way to Price Federal Reserve Services," Federal Reserve Bank of Minneapolis, *Quarterly Review, 1* (Summer 1977), 15-22.

National Commission on Electronic Fund Transfers, *EFT in the United States*. Washington, D.C. (October 1977).

Oi, W.Y., "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly," *Quarterly Journal of Economics, 85* (February 1971), 77-96.

Paulus, John, "The Burden of Federal Reserve Membership, NOW Accounts, and the Payment of Interest on Reserves," *Staff Study*, Board of Governors of the Federal Reserve System, Washington, D.C. (June 1977).

Rhoades, Stephen A., "The Role of Mergers in the Growth of the 20 Largest Banks and Industrials," *Antitrust Bulletin* (forthcoming).

Robert Knight Associates, *Account Analysis Charges for Selected Correspondent Banking Services*. Alliance, Nebraska (1981).

Roberts, K.W.S., "Welfare Considerations of Nonlinear Pricing," *Economic Journal, 89* (March 1979), 66-83.

Rose, John T., "Federal Reserve System Attrition Since 1960," *Journal of Bank Research, 10* (Spring 1979), 8-27.

———, and Peter S. Rose, "The Burden of Federal Reserve System Membership: A Review of the Evidence," *Journal of Banking and Finance, 3* (1979), 331-45.

Sherman, Roger, and Anthony George, "Second-Best Pricing for the U.S. Postal Service," *Southern Economic Journal, 45* (January 1979), 685-95.

Starleaf, Dennis R., "Nonmember Banks and Monetary Control," *Journal of Finance, 30* (September 1975), 995-75.

Survey Research Center, *Household Mail Stream Study* (prepared for the Mail Classification Research Division, Rates and Classification Department, U.S. Postal Service), University of Michigan (December 1978).

———, *Nonhousehold Mail Stream Study* (prepared for the Mail Classification Research Division, Rates and Classification Department, U.S. Postal Service), University of Michigan (July 1980).

———, *The 1984 Survey of Currency and Transaction Account Usage* (prepared for the Division of Research and Statistics, Board of Governors of the Federal Reserve System), University of Michigan (1984).

U.S. Congress, *Depository Institutions Deregulation and Monetary Control Act of 1980*, S. Rept. 96-640, 96 Cong. 1 Sess., Government Printing Office, Washington, D.C. (1980).

U.S. Congress, House of Representatives, Committee on Government Operations and the Committee on Banking, Finance and Urban Affairs, Joint hearings on *The Role of the Federal Reserve in Check Clearing and the Nation's Payments System*, serial No. 98-36, 98 Cong. 1 Sess., Government Printing Office, Washington, D.C. (1983).

Wolkowitz, Benjamin, and Peter Lloyd-Davies, "Reducing Federal Reserve Float," *Federal Reserve Bulletin, 65* (December 1979), 945-50.

Zimmerman, Gary C., "The Pricing of Federal Reserve Services Under the MCA," Federal Reserve Bank of San Francisco *Economic Review*, (Winter 1981), 22-40.

THE MONOGRAPH SERIES IN FINANCE AND ECONOMICS

The following is a list of the available recent Monographs. The price is $5.00 per copy (except as otherwise noted).

1975— 1. Public and Private Offerings of Public Debt: Changes in the Yield Spread—Michel Fleuriet ($3.00)
2. Companies and Common Stocks as Inflation Hedges—Frank K. Reilly ($3.00)
3. "Managed Floating" As An Interim International Exchange rate Regime, 1973-1975—Samuel I. Katz ($3.00)
4. Improving the Access of Developing Countries to the U.S. Capital Market—Robert G. Hawkins, Walter L. Ness, Jr. and Il SaKong ($3.00)
5. Creating New Financial Instruments for the Aged—Jack M. Guttentag ($3.00)

1976— 1. The Stock Price Performance of Firms Repurchasing Their Own Shares—Kenneth R. Marks ($3.00)
2. The Liberalization of U.S. Capital Outflows: Internation Financial Consequences—Kaj Areskoug ($3.00)
3. Money, Credit, and Expenditure: A Sources and Uses of Funds Approach—James S. Earley. Robert T. Parsons, and Fred A. Thompson ($3.00)
4. The Political Economy of Postwar U.S. Trade Policy—Robert E. Baldwin ($3.00)

1977— 1. A Modest Proposal for a National Securities Market System and Its Governance—Ernest Bloch and Arnold W. Sametz ($3.00)
2. The Economic Effects of Compulsory Patent Licensing—F.M. Scherer ($3.00)
3. Profile of the New York Based Securities Industry—Michael Keenan ($3.00)
4. An Empirical Analysis of the Market for Tax-Exempt Securities: Estimates and Forecasts—Patric H. Hendershott and Timothy W. Koch ($3.00)

1978— 1. Tax Systems and Corporate Financing Policy—R.C. Stapleton and C.M. Burke (out of print)
2. Holding Period Yields and Risk-Primum Curves for Long-Term Marketable Securities: 1910-1976—Robert M. Soldofsky and Dale F. Max ($3.00)
3. The Behavior of Foreign Exchange Markets—A Critical Survey of the Empirical Literature—Steven W. Kohlhagen ($3.00)
4. Financial Effects of Capital Tax Reforms—Marshall E. Blume, Jean Crockett, and Irwin Friend($3.00)

5. An Empirical Study of Mortgage Redlining—George J. Benston, Dan Horsky and H. Martin Weingartner ($3.00)

1979— 1. The Implications of Managed Floating Exchange Rates for U.S. Trade Policy—Isaiah Frank, Charles Pearson and James Riedel ($3.00)

2. Regulation of Securities Markets: An Examination of the Effects of Increased Competition—Hans R. Stoll ($3.00)

3. Municipal Revenue Bond Costs and Bank Underwriting : A Survey of the Evidence—William L. Silber ($3.00)

4. Savings Bonds: Theory and Empirical Evidence—Michael J. Brennan and Eduardo S. Schwartz ($4.00)

1980— 1. Redlining and Public Policy—Jack M. Guttentag and Susan M. Wachter ($4.00)

2. Intraindustry Differences in Conduct and Performance: Viable Strategies in U.S. Manufacturing Industries—Richard E. Caves and Thomas A. Pugel ($4.00)

3. The Forward Market for Loanable Funds at Life Insurance Companies—Lawrence J. Horan ($4.00)

4. Discrimination in Mortgage Lending: A Study of Three Cities—A. Thomas King ($4.00)

1981— 1. Averting International Banking Crises—James W. Dean and Ian H. Giddy ($4.00)

2. The Regulation of Foreign Banks: Policy Formation In Countries Outside the United States—Adrian E. Tschoegl ($4.00)

3. The Analysis of Value and Volatility in Financial Futures—Martin Leibowitz ($4.00)

4. Measuring the Importance of Small Business in the American Economy—Lawrence J. White ($4.00)

1982— 1. Toward A Positive Theory of Insurance—David Mayers and Clifford W. Smith, Jr. ($4.00)

2. Do Stock Prices Reflect the Adjustment of Earnings for Inflation?—Phillip Cagan ($4.00)

3-4. Risk Arbitrage II—Guy P. Wyser-Pratte ($10)

1983— 1. Market, Firm and Economic Performance—Stephen Martin ($4.00)

2. Exchange-Rate Regime Selection In Theory and Practice—Gordon Weil ($4.00)

3. The Financial Impacts of Social Security—R. Glenn Hubbard ($4.00)

4. American Banks in the International Market—Stephen V. O. Clarke

1984— 1-2. The U.S. Payments System: Costs, Pricing, Competition and Risk—David Burras Humphrey ($7.50)

SUBSCRIPTIONS

4 to 5 numbers issued per academic year-$15.00
(Special rate to full-time teachers-$7.50 per year)

Subscriptions and orders for individual Monographs should be sent to
Salomon Brothers Center for the Study of Financial Institutions,
90 Trinity Place, New York, New York 10006

Out-of-print MONOGRAPHS and BULLETINS are available from:

University Microfilms International
Library Services
300 North Zeeb Road
Ann Arbor, Michigan 48106, U.S.A.

University Microfilms International
St. Johns Road, Tylers Green
High Wycombe
Buckinghamshire HP 10 8 HR, England

Inquiries should be made to the above addresses.

Customers in JAPAN may find it more convenient to order THE MONOGRAPH SERIES IN FINANCE AND ECONOMICS from the Far Eastern Book-Sellers, No. 2, 2-Chome, Kanda-Jinbocho, Chiyoda-Ku, Tokyo. This bookseller will usually have Salomon Brothers Center's Monographs in stock.

New York University
A private university in the public service

Graduate School of Business Administration

New York University is an affirmative action/equal opportunity institution.

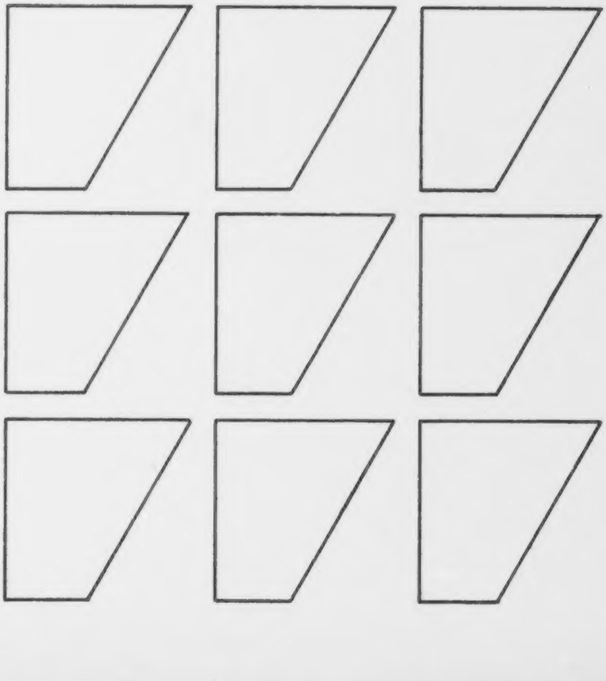
6/11/85 U3515-2.7-F

Monograph Series in Finance and Economics

Monograph 1984-3

The Role of Gold in Consumer Investment Portfolios

By Lawrence S. Ritter and
Thomas J. Urich



Salomon Brothers Center
for the Study of
Financial Institutions
Graduate School of
Business Administration
New York University

SPONSORING SUBSCRIBERS
TO THE MONOGRAPH SERIES IN FINANCE AND ECONOMICS

Bankers Trust Company
Bernstein-Macaulay, Inc.
The Chase Manhattan Bank, N.A.
Chemical Bank
Citibank, N.A.
Cowen & Company
The Dreyfus Corporation
The Equitable Life Assurance Society of the United States
Exxon Corporation
The First National Bank of Chicago
The Guardian Life Insurance Company of America
Manufacturers Hanover Trust Company
Mellon Bank
Metropolitan Life Insurance Company
Morgan Guaranty Trust Company
Phibro-Salomon, Inc.
The Philadelphia National Bank
The Prudential Insurance Co. of America
Salomon Brothers, Inc.

THE MONOGRAPH SERIES IN FINANCE AND ECONOMICS

Editor: Anthony Saunders

The Monograph Series of the Schools of Business of New York University specializes in the publication of original studies in the fields of finance and economics. It provides an outlet for manuscripts which are too long and too broad in scope for the scholarly journals or of too great length and depth for the business periodicals. The submission of manuscripts in these fields is solicited.

For subscription rates and single copy orders, see inside back cover of Monograph.

The Monograph Series is the successor to the *Bulletin* (1925-1964), which was generously supported by C.J. Devine as one of the activities of the Institute of Finance. The Monograph Series is published by the Salomon Brothers Center for the Study of Financial Institutions at the Graduate School of Business Administration of New York University. Editorial offices are located in 1307 Merrill Hall, 90 Trinity Place.

SALOMON BROTHERS CENTER FOR THE STUDY OF FINANCIAL INSTITUTIONS
Director: Arnold W. Sametz

NEW YORK UNIVERSITY GRADUATE SCHOOL OF BUSINESS ADMINISTRATION
Dean: Richard R. West

Biographical Note

Lawrence S. Ritter is Professor of Finance and Economics at New York University. He is former Managing Editor of the Journal of Finance and a past President of the American Finance Association.

Thomas J. Urich is Associate Professor of Economics and Finance at Baruch College, City University of New York. He received his Ph.D. from New York University. Professor Urich has published numerous articles in professional journals in the area of financial markets and investments.

Acknowledgements

Financial support for the research underlying this monograph was provided by International Gold Corporation Limited and by Salomon Brothers Center for the Study of Financial Institutions.

Abstract

Since early 1968, the price of gold has fluctuated between \$35 and \$875 a troy ounce. At the end of 1983, the price was \$380 an ounce, a 16.1 percent per annum nominal rate of increase over the period since the beginning of 1968. After correcting for inflation, the "real" rate of return over that period was 9.0 percent per annum. Although gold is risky in itself, under many circumstances the addition of gold to a portfolio can provide diversification that reduces total portfolio risk. We conclude that on average 5 to 10 percent of one's total investment portfolio in gold is warranted and prudent, although the proportion that is appropriate for any particular individual depends on such factors as how one views the future and how much risk one feels comfortable bearing.

Monograph 1984-3

The Role of Gold in Consumer Investment Portfolios

By Lawrence S. Ritter and
Thomas J. Urich

TABLE OF CONTENTS

Introduction 1

I. Background: Gold In Domestic and International Finance 3

- **A. From the Gold Standard to Floating Exchange Rates** 3
- **B. Will the Gold Standard Make a Comeback?** 8

II. Supply, Demand, and the Price of Gold, 1968-1983 13

- **A. Supply and Demand** 13
- **B. The Price of Gold** 14
- **C. The Rate of Return** 19

III. Looking Backward: Risks and Returns on Alternative Investments 25

- **A. Nominal Yields, 1968-1983: Silver, Housing, Treasury Bills, Stocks, and Long-Term Government Bonds** 25
- **B. Relative Riskiness** 32
- Appendix: The Underlying Data 37

IV. Looking Forward: The Role of Gold In a Modern Investment Portfolio 52

- **A. Gold as a Hedge Against Inflation** 52
- **B. Diversification With Gold Can Reduce Risk** 54
- **C. How Much Gold Should be in a Modern Consumer Portfolio?** 57
- Appendix: Transactions Costs Associated with Different Forms of Gold 62

V. Summary 65

Copyright 1984 by New York University

Introduction

Historically, gold has played an important role in business and government due to its multiple uses as monetary metal, industrial commodity, investment medium, and store of value. However, it has only been since the mid-1970s that individuals in the United States have focused on gold as an investment vehicle. This is because it was not until 1968 that central banks stopped fixing the private transactions price of gold through official cartel-type agreements. (Official transactions prices were maintained until 1973.) And, of course, it was not until December 31, 1974, that individuals in the United States were permitted to own gold.

For anyone who is interested in gold as an investment, it is important to examine how the price of gold is and has been determined in the financial marketplace. It is also important to examine the risk and return characteristics of gold in an investment portfolio and its future possible role. To this end, the present monograph is organized into five chapters.

Chapter I presents an historical overview of the role of gold in domestic and international finance so that we can appreciate its evolution as a monetary metal. Under the full domestic and international gold standard, which existed in one form or another from the late 1800s until the early 1930s, each nation used to set an official price for gold and then stand ready to buy or sell unlimited amounts at that price. The changing monetary role of gold over the following years until 1968, when the private price of gold was no longer fixed by central bank agreement, provides useful insights into short- and long-term fluctuations in gold prices under various monetary systems. The chapter discusses how gold prices reacted under different international monetary arrangements, what role we see for gold as a monetary metal in the future, and the implications for gold prices in the coming years.

Ultimately, the price of any commodity is determined by demand and supply. For gold, supply is relatively fixed while demand fluctuates in response to political and economic conditions. In Chapter II, the demand for gold and the resulting price is examined in light of political and economic developments during the period 1968 through 1983.

In Chapter III, we examine returns on five other consumer investment assets: silver, housing, Treasury bills, long-term Treasury bonds, and common stocks. The risk and return characteristics of these assets are then compared with those of gold.

Chapter IV analyses gold in a modern investment portfolio context. First, the value of gold as a hedge against inflation is compared to the five alternative assets discussed in Chapter III. Next, we explore the risk and return of portfolios with and without gold—that is, we investigate the ability of gold to reduce the overall risk of a portfolio through diversification without equally offsetting reductions in the portfolio overall return. In an Appendix to the chapter, we present data on the transactions costs of investing in gold bullion coin as distinguished from investing in gold bars; the results presented in the monograph are essentially unaffected by the form of gold purchased.

Finally, in Chapter V, we present a summary of our findings.

I. Background: Gold in Domestic and International Finance

A. From the Gold Standard to Floating Exchange Rates

Since gold is historically a monetary metal as well as an industrial commodity, an investment medium, and a store of value, some understanding of its monetary role is essential in analyzing both short- and long-run fluctuations in its price. Under the traditional full domestic and international gold standard—which has not existed since the early thirties—each nation used to set an official price for gold and then stand ready to buy or sell unlimited amounts at that price.

The Gold Standard Act of 1900, for instance, established \$20.67 a troy ounce as the official price of gold. Aside from small handling charges, the United States Treasury would buy any quantity offered at that price, so that no seller had to accept less, and would sell any quantity demanded at that price, so that no buyer had to pay more. As a result, the market price became indistinguishable from the official price, \$20.67 a troy ounce.[1]

In March of 1933, in the depths of the Great Depression, Treasury gold sales were suspended and the United States went off the full gold standard. In January of the following year the U.S. returned to a limited form of international gold standard, with a new official gold price of \$35 an ounce. However, gold coin and bullion were no longer allowed to circulate domestically: American residents were forbidden to buy or hold gold, except for use in industry and the arts or for numismatic purposes, a ban that remained in effect until the end of 1974. The Treasury did business at the new \$35 price only with foreign governments and official monetary institutions.

At \$35 an ounce, gold flowed from abroad into the U.S. Treasury's coffers throughout the 1930s and 1940s. The Treasury's gold stock increased from 290 million ounces in 1935 to 700 million at the end of 1949. The 1950s, however, saw a gradual decline in U.S. gold holdings, a decline that turned into a hemorrhage in the 1960s. By 1969 only about 340 million ounces remained, less than half of what the Treasury had owned in 1949. By the end of 1972, the U.S. gold stock had fallen to 275 million ounces, which was *less* than the amount held in 1935.

---

[1] Gold's weight is customarily measured in troy ounces, which are heavier than the more familiar avoirdupois ounce: one troy ounce = 480 grains while one avoirdupois ounce = 437.5 grains. Thus 1 troy ounce = 1.097 avoirdupois ounces. A metric ton = 32,150 troy ounces.

The inflow of gold to the U.S. during the thirties and forties was at first due to an initial overvaluation of gold, when its price was raised from $20.67 to $35 an ounce early in 1934. The higher price stimulated both production of newly-mined gold and shipment of both existing and newly-mined metal to the United States. Later, the growth of international tensions and then the outbreak of World War II led to a flight of capital from Europe and Asia and a sharp increase in the rate of gold inflow.

The gold outflow from the U.S. during the fifties stemmed from growing U.S. balance of payments deficits, which supplied dollars to the rest of the world in excess of the demand for dollars to purchase U.S. exports. Many of these foreign-held dollars were kept as international monetary reserves by the nations acquiring them, but a large portion was returned through official channels to the United States Treasury with a request that they be exchanged for gold. Although Americans were forbidden to have gold in their possession, the Treasury maintained full convertibility between dollars and gold for foreign governments and central banks. Indeed, this promise of unlimited conversion of dollars into gold on demand was the linch-pin of the international monetary system in the quarter century immediately following World War II.

By 1960, however, the system was already starting to fall apart. America's short-term dollar liabilities to foreigners had grown larger than its shrinking gold holdings, stimulating rumors that soon the United States might once again raise the official price of gold, just as it had in 1934. Raising the official price of gold—that is, devaluing the dollar—would automatically enlarge the dollar value of America's dwindling gold stock.[2]

In late 1960, in anticipation of gold perhaps soon being worth more than $35 an ounce, a heavy private demand drove its market price up to $40 an ounce in London and Zurich, at that time the location of the world's major free gold markets. Free market gold priced at $40 an ounce threatened the gold reserves of the United States as well as of any other nation willing to sell at the $35 official price. Some foreign central banks were tempted to buy gold through official channels at $35 an ounce and then turn around and sell it to private owners in London or Zurich for $40.

The crisis was resolved early in 1961 when eight of the world's leading financial nations banded together and formally established a "gold pool" to peg the market price at $35. The central banks of the United States, Great Britain, Belgium, Holland, West Germany, Italy, France, and Switzerland began to act in concert, as a syndicate, to sell gold to the free market whenever the price began to rise above $35 and to buy it whenever the price started to fall below.

---

[2] Raising the official price of gold and devaluing the dollar are one and the same thing. If $35 = 1 troy ounce of gold, it follows that $1 = 1/35th of an ounce of gold. Raising the price of gold to $40 an ounce means that the dollar has been devalued — i.e., each dollar is now worth less gold (namely, 1/40th of an ounce).

For seven years the gold pool successfully kept the market price within a few cents of $35 an ounce, at the sacrifice of a considerable loss of gold from central bank reserves. The pool had to keep selling, and hardly ever buying, for the same reason it had been organized to begin with: persistent American balance of payments deficits, feeding dollars abroad, continued to stimulate the same expectations about devaluation of the dollar—an increase in the official price of gold—that had driven up the market price of gold in the first place.

Under the circumstances, it was only a matter of time before the next crisis. It arrived late in 1967, as the war in Southeast Asia heated up and accelerated the flow of U.S. dollars abroad. The private demand for gold was so great that by year's end the central banks in the gold pool had almost $2 billion less gold in their vaults than they had at the start of the year. Seeing no end to the drain, the eight gold pool nations ended the arrangement in March of 1968.

The gold pool was replaced by a "two tier" gold price system. In one tier, the official tier, the leading central banks agreed to buy and sell gold only among themselves, with all transactions taking place at the $35-an-ounce official price. They agreed not to buy any newly-mined gold and not to buy or sell any more gold on private markets. The other tier was represented by the private market, where all newly-mined gold had to be channeled and where the price was free to vary in response to changes in supply and demand.

Thus from March 1968 until November 1973, when the two-tier system fell apart, there were two gold prices: the official price and the free market price. The official price was the price at which all governmental and International Monetary Fund settlements were recorded; this was $35 an ounce until it was increased to $38 in December 1971, and then to $42.22 in February of 1973.

The free market price, however, soared to well over $100 an ounce during this five and one-half year period, reaching the then-record high of $125 in mid-1973. So long as the two tiers did not deal with each other, or with intermediaries, there was no necessary connection between the two prices.

Although the two-tier system worked from early 1968 to late 1973, it contained within itself the seeds of its own destruction: the higher the free market price rose above the decreed official price, the greater became the incentive for those governments with relatively large holdings of gold to push for worldwide devaluation up to the market price. Or, alternatively, to surreptitiously sell some of their gold on the private market.

As a result of such pressures, the two-tier gold price system collapsed in November of 1973. It had become too difficult to keep the official and private gold markets in separate watertight compartments. The leading central banks stopped acting in concert to hold down the market price of gold and agreed that any who wished were free to buy or sell gold on their own in the lucrative private market.

The postwar international monetary system established at Bretton Woods, New Hampshire, in 1944 was based on the foundation of fixed exchange rates backed by the ultimate convertibility of the dollar into gold at $35 an ounce. The system finally collapsed in 1971 under the weight of persistent U.S. balance of payments deficits which continued to fuel fears that the United States would have to devalue the dollar.

The possibility that the United States might devalue, decreasing the value of the dollar relative to other currencies, led to panic selling of the dollar in the early months of 1971. To maintain fixed exchange rates, as required by the Bretton Woods Agreement, the central banks of "strong money" countries—such as West Germany, Switzerland, and Japan—had to buy up the flood of dollars being offered on the market, for that purpose paying out their own money. Such purchases, like any open market buying by a central bank, laid the basis for an expansion of their domestic money supplies—in most cases contrary to the prevailing monetary policies of the countries involved. In the process of maintain the *external* value of their money, they were losing control over its *internal* value.

West Germany was the first to abandon ship. In early May of 1971, after buying up several billions of dollars in the first few days of the month, West Germany ceased its dollar purchases, stopped supporting fixed exchange rates, and let the mark float to find its own level. Holland did the same, and for a couple of months the crisis seemed to have been weathered successfully.

In early August, however, there was again a renewed flight from dollars. Faced with jittery and panic-prone international financial markets, with $35 billion in the hands of foreign central banks, obligated to pay out gold to any central bank that wanted to exchange its U.S. dollars for it—but with less than $10 billion worth of gold (valued at $35 an ounce) left in the Treasury—on August 15, 1971, President Richard M. Nixon ended the convertibility of the dollar into gold. Thus the United States went off the international gold standard by executive decree, just as it had gone off the domestic gold standard 38 years earlier, in 1933.

Actually, the termination of the dollar's convertibility into gold caused little disarray in international finance. Everyone had known for the better part of a decade that the commitment of the United States to pay out gold in exchange for dollars was more symbolic than real. If it had ever been put to a real test, it would have ended years before.

Most major countries joined West Germany and Holland and permitted exchange rates to float, with each country's money finding its own value relative to the others on the basis of supply and demand conditions. In mid-December, however, the leading nations resurrected fixed exchange rates as part of a "new look" in international financial arrangements, known as the Smithsonian Agreement. At the same time, the dollar was finally devalued, for the first time since 1934, from the equivalent of 1/35th of an ounce of gold to 1/38th of an ounce. This raised the official price of gold from $35 to $38 an ounce.

The Smithsonian Agreement of December 1971 was hailed by President Nixon as "the most significant monetary agreement in the history of the world." Nevertheless, by February 1973, only fourteen months later, it had almost completely disappeared. At that time the United States again devalued, this time raising the official price of gold an additional 11 percent, to $42.22 an ounce.

The 11 percent devaluation of February 1973 halted panic selling of the dollar only temporarily. In anticipation of *further* devaluation, it soon started again. Finally, in the so-called Brussels Agreement of March 12, 1973, Germany and the major Western European nations decided they had no choice but to let their currencies float relative to the dollar, thereby abandoning fixed exchange rates.

Since March of 1973 the international monetary system has been characterized mainly by floating exchange rates. Several Western European nations fix the value of their monies relative to each other, permitting only limited fluctuations above or below par. And many small less-developed countries peg the international exchange value of their money to that of a larger country, typically a major trading partner. But by and large floating rates predominate, with demand and supply continuously moving foreign exchange rates up and down.

Strictly speaking, floating exchange rates existed "illegally" for almost three years, since the International Monetary Fund's regulations still required fixed rates until early in 1976. Eventually, though, the rules of the game were officially changed at the IMF's meetings in Kingston, Jamaica, in January 1976. At those meetings the entire structure of mandatory fixed exchange rates based on the convertibility of the dollar into gold, erected at Bretton Woods near the close of World War II, was formally dismantled and a new international monetary system was constructed in its place, one that permits flexible exchange rates, although it does not require them.

The Kingston Agreement of 1976 also "demonetized" the international monetary role of gold by legally abolishing the concept of an "official" gold price. Thus the last remnant of the international gold standard was eliminated.

The last remnant of the gold standard *domestically* had been eliminated several years earlier. In the 1950s and early 1960s, the Federal Reserve had been required to hold reserves of 25 percent in the form of gold certificates, representing gold owned by the Treasury, against its two major liabilities: Federal Reserve Notes (most U.S. currency) and bank deposits at the Federal Reserve Banks (bank reserves).

These gold certificate reserves were considered the ultimate backing behind the money supply of the United States, both currency and demand deposits. Indeed, economics textbooks were fond of portraying the domestic monetary system as an inverted pyramid, with a base of gold that supported an equal volume of gold certificates, above that a fourfold expansion of currency and member bank reserves, and atop that a further multiple expansion of demand deposits.

But then, as the Treasury's gold stock shrank from over $20 billion in 1958 to less than $16 billion at the end of 1964, the gold certificate backing behind member bank reserves was legislated out of existence in March of 1965. Three years later, in March of 1968, the corresponding gold certificate reserve behind Federal Reserve Notes was also abolished. Since then, there has been no legal relationship of any sort between gold and the money supply.

Basically, the gold standard has not been in existence domestically since 1933 or internationally since 1971, although it took until 1968 before all domestic vestigial remnants were eliminated and until 1976 before the same was true internationally.

B. Will the Gold Standard Make a Comeback?

The gold standard still retains considerable appeal for many, who periodically urge that it be reinstated. The gold standard is particularly attractive today because it contains built-in safeguards against inflation, worldwide perhaps the most intractible problem of the postwar period. It does this by rigidly tying the money supply to gold, making it difficult if not impossible to enlarge the money supply enough to support any sustained increase in the price level. New discoveries of gold could conceivably produce a dramatic increase in the money supply, but barring that possibility the supply of money is held in check by the existing gold stock plus inflows from abroad and more intensive mining of known sources.

This mandatory link between nation's gold stock and its money supply is intended to supplant human judgment with respect to money. In the absence of such a tether, we are at the mercy of politicians and central bankers, few of whom have been notably successful in curbing inflation.

The tie between gold and the money supply is accomplished by (1) imposing a fixed ratio between gold held by the government and currency in circulation (for every $1 of gold, for example, $4 of currency); and (2) permitting unlimited convertibility between gold and currency. Since bank deposits are interchangeable with currency, in effect deposits are also freely convertible into gold.

Supplementary provisions are often added, such as the use of gold coins as part of the money supply and the requirement that currency be backed 100 percent by gold—that is, that the government hold gold dollar for dollar behind all forms of currency. However, these provisions are not essential. Gold coins are a convenience as part of a gold standard but not a necessity, and they do not by themselves mean that a nation has adopted the gold standard. Both Canada and Mexico currently issue gold coins, but neither is on a gold standard.

And while a gold standard does require that gold be held by the government in some stipulated ratio behind its currency, that ratio need not be 100 percent. Fractional gold reserves are based on the same logic as fractional bank reserves: everyone is unlikely to want to turn their money into gold at one and the same time, just as they are unlikely to want to turn their demand deposits into currency simultaneously.

For conversion between gold and currency, an official fixed price of gold has to be legally established, such as $35 = 1 troy ounce of gold. Since this automatically means that $1 is worth 1/35th of an ounce of gold, it is often expressed by saying that a country thereby "defines" its monetary unit in terms of a specified physical amount of gold: the dollar is defined as equal to 1/35th of an ounce of gold.

Neglecting small handling charges, the public would then be free to convert $35 into one ounce of gold at the Treasury whenever they wished, or to turn in one ounce of gold and get $35. Of course, the Treasury has to convince people that it has enough gold to make good on its promises.

This convertibility between currency and gold is the lever that controls the money supply, and through the money supply presumably the price level. This is because many people are convinced, on the basis of experience, that over the long run gold will retain its value better than paper money. Expectations that prices are apt to rise is thus likely to lead the public to exchange currency for gold. And, since bank deposits are freely convertible into currency, to also exchange deposits for currency and then for gold. Inflationary expectations thereby generate their own remedy automatically, without intervention by politicians or central bankers, since the smaller money supply aborts inflation before it can really get underway.[3]

If the gold standard functions so smoothly and contains inflation so effectively, why was it abandoned by the United States (and the rest of the world as well)?

Although the gold standard has built-in safeguards against inflation, the same cannot be said with respect to recession and unemployment. Under the gold standard, the money supply is determined by the public, not by the central bank. Gold hoarding by the public shrinks the money supply and the central bank, if it abides by the rules, cannot do anything about it. This is effective in stopping inflation but is likely to *intensify* a recession, especially if fears arise that the government might possibly devalue (that is, raise the price of gold).

Governments are often tempted to devalue during recessions, in order to increase employment by expanding exports. Devaluation depreciates the foreign exchange value of a country's money, thereby inhibiting its imports and stimulating its exports.

If the public starts to suspect that the government might soon devalue, it makes sense to buy gold now, before the price rises, and then sell it back to the government after the price has gone up. This is precisely what led the United Kingdom and the United States to leave the gold standard fifty years ago. The United Kingdom suspended convertibility of the pound into gold, and thereby went off the gold standard, in September of 1931. The United States followed suit a year and a half later. The reasons were the same in both cases: large gold drains, due to unsettled world economic conditions and persistent rumors of imminent devaluation despite repeated government denials that anything of the sort was contemplated.

Under the gold standard, a country losing gold is supposed to contract its money supply and put deliberate downward pressure on its economy. In the early thirties, both Great Britain and the United States were indeed losing gold, but understandably neither wanted to impose the tight money called for by the gold standard's rules. Both were in the midst of major depressions that neither wanted to exacerbate by monetary stringency. The generally accepted remedy for a depression is easier, not tighter, money.

---

[3] There are international self-correcting counterparts that are similar to these domestic forces—balance of payments deficits (and surpluses) that generate gold outflows (and inflows).

Nor did either know, at the time, what new official price to set for gold that would stem the outflow. If they set too high a price, other nations would retaliate and devalue as well; if they set too low a price, the gold outflow would continue. If too much gold was lost, they would be forced off the gold standard anyway. It was easier to simply suspend convertibility and wait to see what happened.

Similar circumstances led to the final breakdown of what was left of the gold standard in the early seventies. The United States resumed convertibility of the (devalued) dollar in 1934, for foreign central banks and governments only. But the system collapsed again in 1971 under the weight of persistent U.S. balance of payments deficits, unstable world financial conditions, and recurrent rumors that the U.S. was again ready to devalue the dollar.

All of which is history. Looking to the future, what are the probabilities that we will return to the gold standard in the foreseeable future? Extremely low, for a number of reasons:

1. It is difficult today, perhaps impossible, to establish a fixed relationship between a nation's gold stock and its money supply. The gold standard was the construct of a simpler monetary era. Most of our money now consist of bank deposits, not currency, and there is even a great deal of controversy about which deposits to include. As a result, we have several different money supplies—M1, M2, M3, and so on—and it is hard to justify why one and not the other should be tied to gold.

2. Just as difficult is the choice of an appropriate official price for gold, given the wide swings in its free market price in recent years. Too low an official price would encourage large scale gold flows from government to private hands, too high a price the reverse.

3. The gold standard has built-in safeguards against inflation but not against recession and unemployment. No democratically elected government could deliberately provoke heavy unemployment, simply because of gold outflows, and expect to remain in office. Indeed, even prior to the 1930s the U.S. "sterilized" the domestic monetary effects of international gold flows by offsetting central bank open market operations—a clear violation of the gold standard's

rules—in order to pursue domestic stabilization objectives unhampered by gold constraints.[4]

For investors in gold who expect that the United States *will* return to the gold standard in the foreseeable future, the most important factor determining the profitability of their investment is what new official price will be established for gold. Since the market price will converge with the official price, only the official price will be relevant. A "high" official price will produce profits while a "low" one will mean losses.

The authors of this monograph assume, however, for the reasons stated above, that the gold standard will *not* be restored in the foreseeable future. Thus there will be no "official" price but rather a free market price determined by supply and demand. What has happened to the market price in the past is relevant information for anticipating what may happen to it in the future. The following chapter therefore analyzes how the market price of gold has behaved, and the forces that have affected it, in the sixteen years since the central bank gold pool was terminated in March of 1968 and the price of gold was allowed to find its own level.

---

[4] A recent Presidential Commission appointed to study the advisability of returning to the gold standard recommended against it. See the *Report to the Congress of the Commission on the Role of Gold in the Domestic and International Monetary Systems* (U.S. Government Printing Office, March 1982).

II. Supply, Demand, and the Price of Gold, 1968-1983

A. Supply and Demand

It has been estimated that about 2.9 billion troy ounces of gold (roughly 90,000 metric tons) has been extracted from the earth from the beginning of recorded history to the present. Around 300 million ounces, close to 10 percent of it, has been lost in one way or another[5]. Much of the remaining 2.6 billion ounces, about 1.2 billion ounces of it, is held as monetary reserves by various governments and official international organizations. The United States Treasury owns 260 million ounces, international organizations like the International Monetary Fund own 200 million ounces, other non-Communist governments and central banks own 670 million, and the Soviet Union and China possess about 70 million (mostly owned by the Soviet Union).

However, the largest share, about 1.4 billion ounces, remains in private hands. Perhaps half of this has been used for artistic and decorative purposes, including religious ornamentation, and been fabricated into jewelry. The rest, about 700 million ounces, is hoarded in the form of gold coin or bullion: an estimated 200 million ounces is hoarded by private citizens in France, 120 million ounces in the United States, an equal amount in India, with an estimated 260 million ounces scattered throughout the Middle East and the rest of the world.

The annual production of newly-mined gold adds only about 1½ to 2 percent a year to this stock—about 40 to 50 million ounces is mined annually. Approximately half of this usually comes from South Africa, close to 25 percent from the Soviet Union, and slightly less than 5 percent from Canada, the three largest gold producers. In recent years the United States has typically accounted for 2 to 3 percent of new output.

This means that the overwhelming element on the supply side is not the amount of current production, which is only a small portion of the total stock in existence, but uncertainty regarding how much holders of the existing stockpile might decide to sell. For instance, the United States Treasury sold 1.3 million ounces out of its official gold stock in 1975, and another 16 million ounces at regular monthly auctions in 1978 and 1979.

---

[5] Data in this subsequent paragraphs is from J. Aron & Company, *Gold Statistics and Analysis*, annual.

The 1975 sale was a deliberate attempt to stop a rise in the price of gold, which it succeeded in doing temporarily. The 1978-79 sales were intended to prop up the foreign exchange value of the dollar by improving the U.S. balance of trade through lower gold imports (to the extent that the auctioned gold was purchased by domestic bidders) and by increased gold exports (to the extent it was purchased by foreigners).

Similarly, at the January 1976 International Monetary Fund meetings in Kingston, Jamaica, the IMF decided to sell 25 million ounces of its gold holdings on the open market and use the proceeds to aid developing countries. Auctions held from 1976 through 1980 raised $4.6 billion for this purpose.

On the demand side, a variable portion of the demand for gold is not for industrial, dental, or artistic use but rather is motivated by psychological considerations—in particular, by periodic fears regarding an uncertain social and economic future. Traditionally, gold has been considered a hedge against inflation. When consumer prices threaten to get out of hand, the demand for gold usually expands; when inflation subsides, that demand often contracts just as suddenly. Similar reactions are associated with periods of social and political turbulence. If a government falls, its paper money often falls with it; gold is a traditional refuge from political upheaval. Even moderate changes in demand, impinging on a relatively fixed aggregate supply that cannot be meaningfully increased by current production, can produce substantial price movements.

B. The Price of Gold

Table 1 records the free market price of gold from early 1968, when the central banks abandoned fixing the price through the gold pool, to the end of 1983. Six major price trends are discernible in that sixteen-year period:

1. *From early 1968 to the end of 1974, the market price of gold rose from the official figure of $35 an ounce to a high of $197 an ounce*. This seven-year time span encompassed the turmoil associated with the Viet Nam War, the Watergate hearings, the resignation of the vice-president and the president of the United States, the 1973 Arab oil embargo, and the 1973-74 400 percent OPEC increase in the price of oil. Consumer prices in the United States rose by almost 9 percent in 1973, the highest annual rate of inflation in a quarter century, and then by over 12 percent in 1974. The then-record high of $197 an ounce was reached late in December of 1974.

2. *From January 1975 to August 1976, the price of gold fell from almost $200 an ounce to just over $100.* Americans were legally permitted to own gold on the last day of 1974—for the first time since 1933—which allowed them to buy some just in time to see its value fall by about 50 percent over the next twenty months. Part of the reason the price rise ended in early 1975 was the U.S. Treasury's decision to sell off some of its official gold stock. The Treasury's first auction, at which it sold 750,000 ounces, was held on January 6, 1975. More important, political and social stability returned to the United States and the worldwide pace of inflation slackened: after 12.2 percent inflation in the United States in 1974, prices rose by 7 percent in 1975 and less than 5 percent in 1976.

3. *From August 1976 to near the end of January 1980, however, the price of gold skyrocked from $104 an ounce to an all-time high of $875 an ounce.* This three-and-a-half-year period was marked by accelerating worldwide inflation; in the United States, consumer prices rose by almost 7 percent in 1977, by 9 percent in 1978, by 13.3 percent in 1979, and by over 12 percent in 1980. The price of gold on the London market went over $200 an ounce for the first time in July of 1978 and then kept on rising for the next year and a half almost without interruption. This despite additional gold auctions by the U.S. Treasury in 1978 and 1979 and a series of similar sales by the International Monetary Fund that began in 1976 and continued until mid-1980.

Fueled by monetary instability, oil shortages, the Iranian hostage crisis, and fears of runaway inflation, the rise in the price of gold during the second half of 1979 and the first month of 1980 reached unprecedented proportions. The price passed $300 an ounce in July of 1979, $400 in October, and went over $500 on December 26, 1979. With the flames fanned by the Soviet invasion of Afghanistan on December 30, the price jumped past $600 an ounce in early January. From $684 an ounce in London on January 15, it rose spectacularly to $765 on the 16th, to $820 a few days later, and then to the all-time high of $875 in New York on January 21, 1980.

4. *From January 1980 to June 1982, the price of gold retreated irregularly from its $875 an ounce peak to slightly below $300.* Following the January 21 high of $875, the price of gold in New York plummeted to $450 in late March and then rebounded to the $700 range briefly in late summer, when the war between Iran and Iraq began. Thereafter, it declined rather steadily over the next two years to a trough of $296 on June 21, 1982.

TABLE 1

The Price of Gold Bullion, 1968-1983

	1968	1969	1970	1971	1972	1973	1974	1975
Jan.	$35.20	$42.60	$35.20	$38.35	$47.45	$66.30	$132.80	$176.05
Feb.	35.20	42.95	35.20	39.10	48.50	85.30	162.80	182.00
Mar.	38.85	43.15	35.50	39.15	48.65	90.30	173.30	177.25
Apr.	39.45	43.80	36.10	39.95	49.90	91.00	169.55	167.00
May	42.20	43.45	35.85	41.15	59.75	115.05	157.05	167.20
June	41.05	41.45	35.75	40.40	64.95	123.55	144.55	166.35
July	38.90	41.60	35.55	42.70	68.60	115.90	156.25	166.95
Aug.	39.90	41.10	36.10	41.00	67.15	103.80	156.30	160.00
Sept.	39.60	40.65	36.65	42.90	64.50	100.30	151.55	141.40
Oct.	39.15	40.00	37.55	42.65	64.65	98.30	167.30	143.05
Nov.	40.05	35.80	37.80	43.95	63.90	101.25	184.55	140.70
Dec.	42.05	35.45	37.65	43.85	65.20	112.25	186.75	140.35

TABLE 1

The Price of Gold Bullion, 1968-1983

	1976	1977	1978	1979	1980	1981	1982	1983
Jan.	$128.40	$132.30	$176.10	$233.70	$653.00	$506.50	$387.00	499.50
Feb.	132.55	142.75	182.25	251.55	637.00	489.00	362.60	408.50
Mar.	129.60	148.90	181.60	240.35	495.00	513.75	320.00	411.75
Apr.	128.40	147.25	170.85	245.55	518.00	482.75	361.25	429.25
May	125.50	143.30	184.15	274.85	535.50	479.25	325.25	437.50
June	123.80	143.00	183.05	277.50	653.50	426.00	317.50	416.00
July	112.80	144.35	200.50	296.70	614.25	406.00	342.90	422.00
Aug.	104.35	146.00	208.70	315.35	631.25	427.00	411.50	414.25
Sept.	116.25	154.20	217.10	397.50	666.75	428.75	397.60	405.00
Oct.	123.15	161.50	242.60	382.25	629.00	427.00	423.25	382.00
Nov.	130.25	160.05	193.40	415.90	623.00	414.50	436.00	405.00
Dec.	134.75	164.95	226.00	512.00	586.00	401.00	456.90	381.50

Source: Handy and Harman. Morning price (10:30 A.M.) on last business day of month.

The extent of the rise in the price of gold in 1979 and early 1980 probably made a downward adjustment inevitable. However, it was aided and abetted by a change in the conduct of Federal Reserve policy, announced in October of 1979: the central bank shifted from preoccupation with interest rates to an emphasis on control over bank reserves and the money supply, letting interest rates take care of themselves. As the new Fed policy took hold in 1980 and 1981, it generated historic increases in interest rates and broke the back of the inflationary spiral.

The discount rate was raised to 14 percent plus a 4 percent surcharge for frequent borrowers, 3-month Treasury bill bond-equivalent yields exceeded 17 percent, and the commercial bank prime loan rate hit 21½ percent, in each case an all-time high. Unemployment in the United States swelled to double digits while inflation subsided—from 12.4 percent in 1980 to less than 9 percent in 1981 and less than 4 percent in 1983.

The historically high interest rates that prevailed during much of 1980 and 1981 played a pivotal role in bringing about lower gold prices. In the late seventies, "real" interest rates—market interest rates minus the rate of inflation—had actually been negative. That is, the rate of inflation was often higher than market yields on financial assets. As a result, investors in financial instruments frequently ended up with less purchasing power than they started with. Under the circumstances, they sought alternatives, including gold. While gold yields no interest or dividend income, and may even entail carrying costs, the conventional expectation is that it will rise in price faster than the rate of inflation.

The high positive real interest rates of the early eighties, with market yields well above the inflation rate, enhanced the attractiveness of financial assets relative to non-interest bearing assets, such as gold, silver, diamonds, numismatic coins, rare stamps, and so on. The result was an expansion in the demand for financial assets and a corresponding contraction in the private investment demand for gold and other so-called "collectibles". The worldwide private investment acquisition of gold coins, medallions, and bullion, which had totalled over 22 million ounces in 1979, thus shrank to about 10 million ounces in 1981.

High positive real interest also inhibited business spending on plant and equipment as well as purchases of new homes, both of which helped dampen economic expansion and subdue inflation. Just as the expectation of more inflation typically expands the demand for gold and drives its price up, so the anticipation of less inflation generally contracts demand and pushes the price down.

5. *From June 1982 to the close of the year, the price of gold see-sawed on an upward trend from $296 an ounce to end the year fifty percent higher at about $450 an ounce.* The rise in the price in the second half of 1982 stemmed from an expansion in the demand for gold generally attributed to a renewal of concern over inflation as well as apprehension about the solvency of financial institutions worldwide. At the time, the international financial system was feeling the initial strains associated with the nonrepayment of debt by many third-world and other debtor countries.

6. *From January 1983 to December 1983, the price of gold continued to see-saw, but now on a downward trend, from a peak of about $500 an ounce in January to about $380 an ounce at year-end.* The 1983 period may be generally characterized as once of increased optimism. There was a lowering of inflationary expectations and new optimism about the course of economic activity which can be seen by the high value of the dollar in international markets and the strong performance of the stock market—the return on equities was over 20 percent for the second year in a row. While there was continued concern about the stability of the international financial system, the "good" news about inflation and growth acted as a damper on gold prices.

High positive real interest rates also strongly contributed to the downward pressure on the price of gold. The demand for financial assets expanded as the real rate of return on financial assets remained high, while there was a corresponding shrinkage in the demand for gold as the real opportunity cost of holding it increased. The overall result was a reduction of over $100 in the price of gold from January to December.

C. The Rate of Return

Overall, from the beginning of 1968, when the price of gold was allowed to find its free market level, to the end of 1983, the market price of gold rose from $35 to about $380 an ounce. Over the full 16-year period, this amounts to a 16.1 percent per annum nominal rate of increase.

Table 2, constructed on the basis of the data in Table 1, shows the annual rate of return on gold for the full period, from the start of 1968 to the end of 1983, as well as rates of return for each yearly subperiod. The table shows the per annum rate of return accruing to someone who acquires gold at the start of any year and holds it to the end of the year or the end of any subsequent year.

For example, Table 2 indicates that a purchaser of gold at the start of 1968, who held it until the end of the year, would have earned a 19.5 percent rate of return. If held for two years, until the end of 1969, the annual yield would have plummeted to 0.4 percent . . . but if held for seven years, until the end of 1974, the annual rate of return would have skyrocketed to 26.9 percent.

Casual observation of Table 2 is sufficient to indicate the importance of timing with respect to investment in gold. A purchase at the start of 1975, for instance, would have produced a loss of 24.9 percent if sold at the end of 1975 . . . but yielded a 22.4 percent per annum gain if held until the end of 1979. Similar illustrations abound in Table 2. We will return to the subject of risk in the next chapter.

Table 3 shows annual rates of inflation on the same basis as Table 2's rates of return. Subtracting the rate of inflation in Table 3 from the nominal rate of return in Table 2 produces the "real" rate of return in Table 4.

From the beginning of 1968 to the end of 1983, the pace of inflation as measured by the consumer price index rose at an annual rate of 7.1 percent. Thus since 1968 the "real" rate of return on gold, not taking carrying costs into account, amounted to 9.0 percent. (The 16.1 percent in Table 2 minus the 7.1 percent in Table 3 equals the 9.0 percent in Table 4.)

Another way to show the implications of the rate of return on gold, unadjusted for inflation, is indicated in Table 5, which contains the amounts to which $1 invested in gold would have grown over different time intervals. For instance, $1 invested in gold at the beginning of 1968 would have grown to $1.19 by the end of 1968 . . . to $5.31 by the end of 1974 . . . to $16.65 by the end of 1980 . . . and to $10.84 by the end of 1983. Growth from $1 in 1968 to $10.84 in 1983 is the equivalent of a 16.1 percent annual rate of return (as shown in Table 2).

What have other investments yielded over that time span? How risky has it been to hold gold? To hold other assets? We turn to these and related questions in the following chapter.

TABLE 2

Rate of Return on Gold, 1968-1983

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	19.5%															
1969	0.4	−15.7%														
1970	2.3	− 5.4	6.2%													
1971	5.7	1.4	11.2	16.5%												
1972	13.1	11.6	22.5	31.6	48.7%											
1973	21.3	21.7	33.4	43.9	60.0	72.2%										
1974	26.9	28.2	39.4	49.2	62.1	69.2	66.4%									
1975	18.9	18.8	25.8	30.1	33.8	29.1	11.8	−24.9%								
1976	16.1	15.7	21.0	23.7	25.2	19.9	6.3	−15.1	− 4.0%							
1977	16.7	16.4	21.2	23.5	24.7	20.4	10.1	− 4.1	8.4	22.4%						
1978	18.4	18.3	22.9	25.1	26.4	23.0	15.0	4.9	17.2	29.5	37.0%					
1979	25.0	25.5	30.6	33.6	36.0	34.2	28.8	22.4	38.2	56.0	76.2	126.6%				
1980	24.2	24.6	29.1	31.6	33.4	31.6	26.6	21.0	33.1	44.4	52.6	61.0	14.5%			
1981	19.0	18.9	22.4	24.0	24.8	22.4	17.3	11.5	19.1	24.4	24.9	21.1	−11.5	−31.6%		
1982	18.6	18.6	21.7	23.1	23.8	21.5	16.9	11.8	18.4	22.6	22.6	19.2	− 3.7	−11.7	13.9%	
1983	16.1	15.8	18.5	19.5	19.8	17.4	13.0	8.3	13.3	16.0	15.0	11.0	− 7.1	−13.3	− 2.5	−16.5%

Source: Table 1.

TABLE 3

The Rate of Inflation, 1968-1983

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	4.7%															
1969	5.4	6.1%														
1970	5.4	5.8	5.5%													
1971	4.9	5.0	4.4	3.4%												
1972	4.6	4.6	4.1	3.4	3.4%											
1973	5.3	5.4	5.2	5.2	6.1	8.8%										
1974	6.3	6.5	6.6	6.9	8.1	10.5	12.2%									
1975	6.4	6.6	6.7	6.9	7.8	9.3	9.6	7.0%								
1976	6.2	6.4	6.4	6.6	7.2	8.2	8.0	5.9	4.8%							
1977	6.2	6.4	6.4	6.6	7.1	7.9	7.7	6.2	5.8	6.8%						
1978	6.5	6.7	6.7	6.9	7.4	8.1	7.9	6.9	6.9	7.9	9.0%					
1979	7.0	7.3	7.4	7.6	8.1	8.8	8.8	8.2	8.4	9.7	11.2	13.3%				
1980	7.4	7.7	7.8	8.1	8.6	9.3	9.3	8.9	9.2	10.4	11.6	12.9	12.4%			
1981	7.6	7.8	7.9	8.1	8.6	9.2	9.3	8.9	9.2	10.1	10.9	11.5	10.7	8.9%		
1982	7.3	7.5	7.6	7.8	8.2	8.7	8.7	8.2	8.4	9.0	9.5	9.6	8.4	6.4	4.2%	
1983	7.1	7.2	7.3	7.5	7.8	8.2	8.2	7.7	7.8	8.3	8.5	8.4	7.2	5.6	4.0	3.8%

The rate of inflation was calculated as the percentage change in the Consumer Price Index.

Source: Bureau of Labor Statistics, U.S. Department of Labor.

TABLE 4

“Real” Rate of Return of Gold, 1968-1983

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	14.8%															
1969	-5.0	-21.8%														
1970	-3.1	-11.2	0.7%													
1971	0.8	- 3.6	6.8	13.1%												
1972	8.5	7.0	18.4	28.2	45.3%											
1973	16.0	16.3	28.2	38.7	53.9	63.4%										
1974	20.6	21.7	32.8	42.3	54.0	58.7	54.2%									
1975	12.5	12.2	19.1	23.2	26.0	19.8	2.2	-31.9%								
1976	9.9	9.3	14.6	17.1	18.0	11.7	-1.7	-21.0	-8.8%							
1977	10.5	10.0	14.8	16.9	17.6	12.5	2.4	-10.3	2.6	15.6%						
1978	11.9	11.6	16.2	18.2	19.0	14.9	7.1	- 2.0	10.3	21.6	28.0%					
1979	18.0	18.2	23.2	26.0	27.9	25.4	20.0	14.2	29.8	46.3	65.0	113.3%				
1980	16.8	16.9	21.3	23.5	24.8	22.3	17.3	12.1	23.9	34.0	41.0	48.1	2.1%			
1981	11.4	11.1	14.5	15.9	16.2	13.2	8.0	2.6	9.9	14.3	14.0	9.6	-22.2	-40.5%		
1982	11.3	11.1	14.1	15.3	15.6	12.8	8.2	3.6	10.0	13.6	13.1	9.6	-12.1	-18.1	9.9%	
1983	9.0	8.6	11.2	12.0	12.0	9.2	4.8	0.6	5.5	7.7	6.5	2.6	-14.3	-18.9	- 6.5	-20.3%

Source: Tables 2 and 3.

TABLE 5

Amount to Which $1 Invested in Gold Would Grow

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	$ 1.19															
1969	1.01	$ 0.84														
1970	1.07	0.90	$ 1.06													
1971	1.25	1.04	1.24	$ 1.16												
1972	1.85	1.55	1.84	1.73	$ 1.49											
1973	3.19	2.67	3.17	2.98	2.56	$1.72										
1974	5.31	4.44	5.27	4.96	4.26	2.86	$1.66									
1975	3.99	3.34	3.96	3.73	3.20	2.15	1.25	$0.75								
1976	3.83	3.20	3.80	3.58	3.07	2.07	1.20	0.72	$0.96							
1977	4.69	3.92	4.65	4.38	3.76	2.53	1.47	0.88	1.18	$1.22						
1978	6.42	5.37	6.38	6.00	5.15	3.47	2.01	1.21	1.61	1.68	$1.37					
1979	14.55	12.18	14.44	13.60	11.68	7.85	4.56	2.74	3.65	3.80	3.10	$2.27				
1980	16.65	13.94	16.53	15.56	13.36	8.99	5.22	3.14	4.18	4.35	3.55	2.59	$1.14			
1981	11.39	9.54	11.31	10.65	9.14	6.15	3.57	2.15	2.86	2.98	2.43	1.77	0.78	$0.68		
1982	12.98	10.87	12.89	12.14	10.42	7.01	4.7	2.45	3.26	3.39	2.77	2.02	0.89	0.78	$1.14	
1983	10.84	9.07	10.76	10.13	8.70	5.85	3.40	2.04	2.72	2.83	2.31	1.69	0.75	0.65	0.95	$0.84

Source: Tables 1 and 2.

III. Looking Backward: Risks and Returns on Alternative Investments

A. Nominal Yields, 1968-1983: Silver, Housing, Treasury Bills, Stocks, and Long-Term Government Bonds

The previous chapter showed that gold yielded a 16.1 percent annual rate of return from 1968 through 1983, increasing in price from $35 to about $380 an ounce. A dollar invested in gold at the beginning of 1968 would have grown to $10.84 by the end of 1983.

In this chapter we examine annual rates of return over the same time interval for five alternative investments: for silver and housing, two alternative non-financial assets, and for Treasury bills, stocks, and long-term government bonds, three alternative financial assets. Chart 1 and Table 1 show the year-by-year cumulative wealth results of investing $1 in gold and in each of the five alternative assets at the end of 1967—in silver, housing, treasury bills, the stock market, and long-term government bonds. Clearly gold, with $1 growing to $10.84 over the entire 16-year time span, would have been the most profitable investment of the six. Silver, stocks, housing, treasury bills, and long-term government bonds follow in that order.

Silver. The price of silver was dominated during this period by the unusual activities of Nelson Bunker Hunt and W. Herbert Hunt of Dallas, Texas. The Hunt brothers amassed silver holdings estimated at over 200 million ounces during 1979 and early 1980 and helped drive up the price of silver to its all-time peak of $50.50 an ounce in mid-January of 1980. Only six months earlier, the price of silver had been below $10 an ounce.

The brokerage firm of Bache Halsey Stuart Shields touched off near panic in the silver market on March 27, 1980, with the announcement that the Hunt brothers had failed to meet $100 million of margin calls on their silver accounts. Bache incurred substantial losses as a result. Subsequently, the Hunts signed an agreement to cover another $400 million in losses that resulted from their silver trading activities. The brothers agreed to turn over oil and gas properties and silver holdings to Engelhard Minerals in exchange for Engelhard's consent to let them out of a contract to buy 19 million ounces of silver at $35 an ounce. The contract came due on March 31st, when silver closed in New York at less than $15 an ounce. As part of the agreement, Engelhard kept its 19 million ounces of silver, received another 9 million ounces from the Hunts, and also received oil and gas properties valued at several hundred million dollars.

Table 1, in the Appendix at the end of this chapter, shows the noon price of silver on the last business day of each month from 1968 through 1983. Also in the Appendix, Table 2 shows the rate of return on investments in silver during that period of time and Table 3 shows the amounts to which $1 invested in silver would grow over various time intervals within the overall time span 1968-1983.

The distortions introduced into the silver market by the apparent attempt of the Hunt brothers to corner the market in 1979 and 1980 make the 1968-83 period atypical of normal or representative movements in the price of silver. Nevertheless, at least for the record, $1 invested in silver at the start of 1968 would have grown to $4.74 at the end of 1983, as indicated by Table 3. This represents a 10.2 percent annual rate of return (see Table 2). These figures fall far short of gold's performance over the same time interval, but greatly exceed the returns yielded by other assets.

TABLE 1

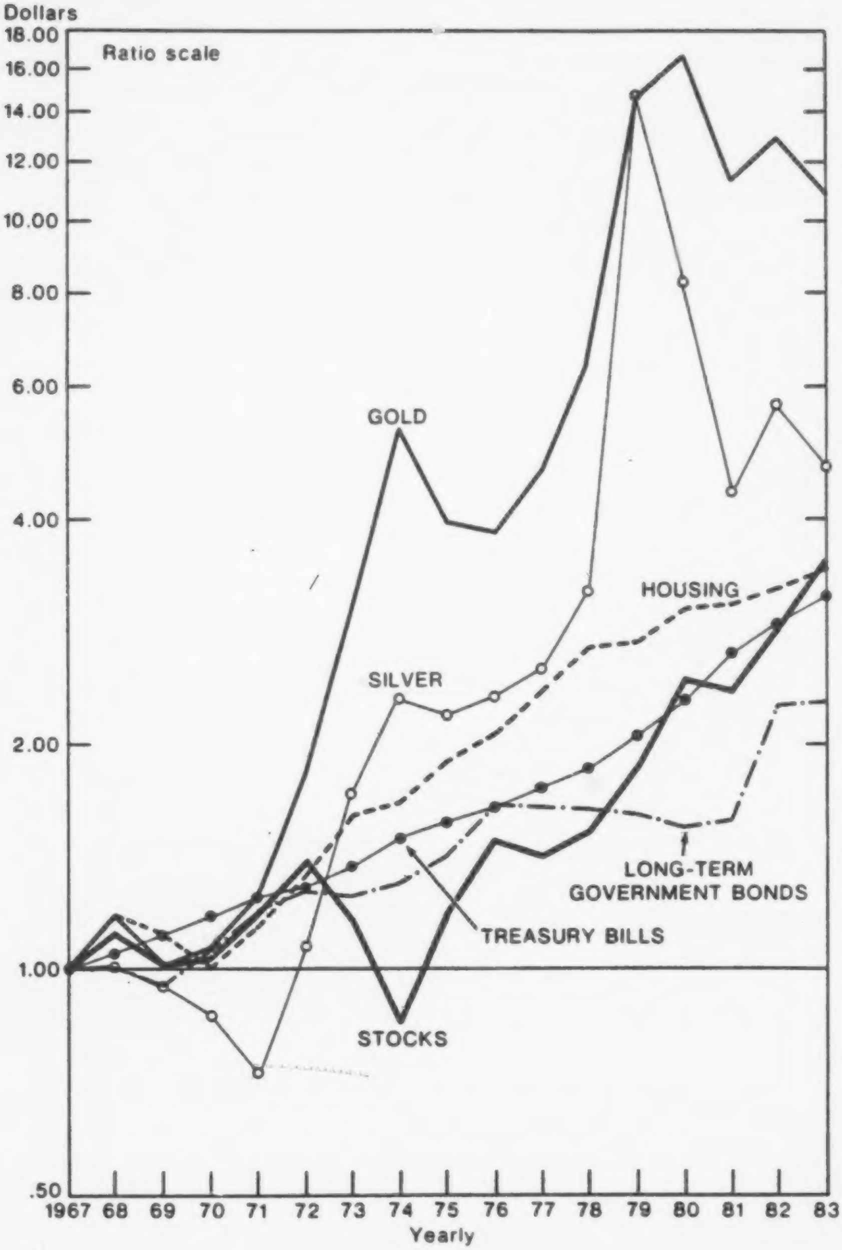
Wealth Indices of Six Alternative Investments

End of	L-T Govt Bonds	Tr. Bills	Housing	Stocks	Silver	Gold
1967	$1.00	$1.00	$1.00	$1.00	$ 1.00	$ 1.00
1968	1.00	1.05	1.19	1.13	1.01	1.19
1969	0.95	1.12	1.13	1.02	0.95	1.01
1970	1.06	1.19	1.00	1.03	0.87	1.07
1971	1.20	1.25	1.14	1.19	0.73	1.25
1972	1.27	1.29	1.34	1.41	1.08	1.85
1973	1.26	1.38	1.61	1.17	1.73	3.19
1974	1.31	1.50	1.68	0.85	2.31	5.31
1975	1.43	1.58	1.90	1.18	2.20	3.99
1976	1.67	1.66	2.07	1.49	2.31	3.83
1977	1.66	1.75	2.37	1.42	2.53	4.69
1978	1.64	1.87	2.70	1.52	3.21	6.42
1979	1.62	2.07	2.77	1.85	14.81	14.55
1980	1.56	2.30	3.02	2.46	8.28	16.65
1981	1.59	2.64	3.08	2.36	4.37	11.39
1982	2.24	2.91	3.23	2.84	5.77	12.98
1983	2.27	3.17	3.42	3.49	4.74	10.84

Source: Table 5 in Chapter II and Tables 3, 6, 8, 10, and 12 in the Appendix to Chapter III.

Chart 1

WEALTH INDICES OF SIX ALTERNATIVE INVESTMENTS



Source: Table I.

Housing. Most Americans did not invest in either gold or silver in the postwar period. Indeed, it was not legal for Americans to buy gold for investment purposes until the beginning of 1975. Thus the best inflation hedge for most Americans during the years 1968-1983 was home ownership. Table 4 in the Appendix to this chapter shows the median sales price of new one-family houses, monthly, 1968-1983, as published by the Department of Housing and Urban Development, Bureau of the Census, U.S. Department of Commerce.

Table 5 in the Appendix shows the annual rate of return on investment in home ownership, based on the median sales price of new one-family houses in Table 4. Table 6 is an alternative way of showing the same thing: it shows the amounts to which $1 invested in housing would grow over various time intervals. For the most part, home prices advanced rather steadily over the years 1968-1983 except for brief dips in 1969-1970 and some declines in the late seventies and early eighties when high mortgage interest rates made it difficult for many potential home buyers to finance purchases.

Table 6 indicates that a dollar invested in housing at the start of 1968 would have grown to $3.42 at the end of 1983. This represents an 8.0 percent annual rate of return. The rate of inflation over this period was 7.1 percent (see Table 3 in Chapter 2), so that on average home buying yielded a positive real rate of return.

Treasury Bills. Our measure of Treasury bill yields consists of one-month holding period returns for one-month bills, following the pioneering work of Ibbotson and Sinquefield.[6] Tables 7 and 8 in the Appendix to this chapter show, respectively, the annual rates of return on Treasury bills for various holding periods over the time span 1968-83, and the amount to which $1 invested in bills would have grown during that period.

As Chart 2 indicates, Treasury bill yields were rather closely related to the inflation rate during most of the sixteen-year time span. In fact, the average annual rate of return on bills and the rate of inflation over the entire period are virtually identical: the annual yield on bills held from the start of 1968 to the end of 1983 was 7.5 percent, while the inflation rate over the same period was 7.1 percent per annum.

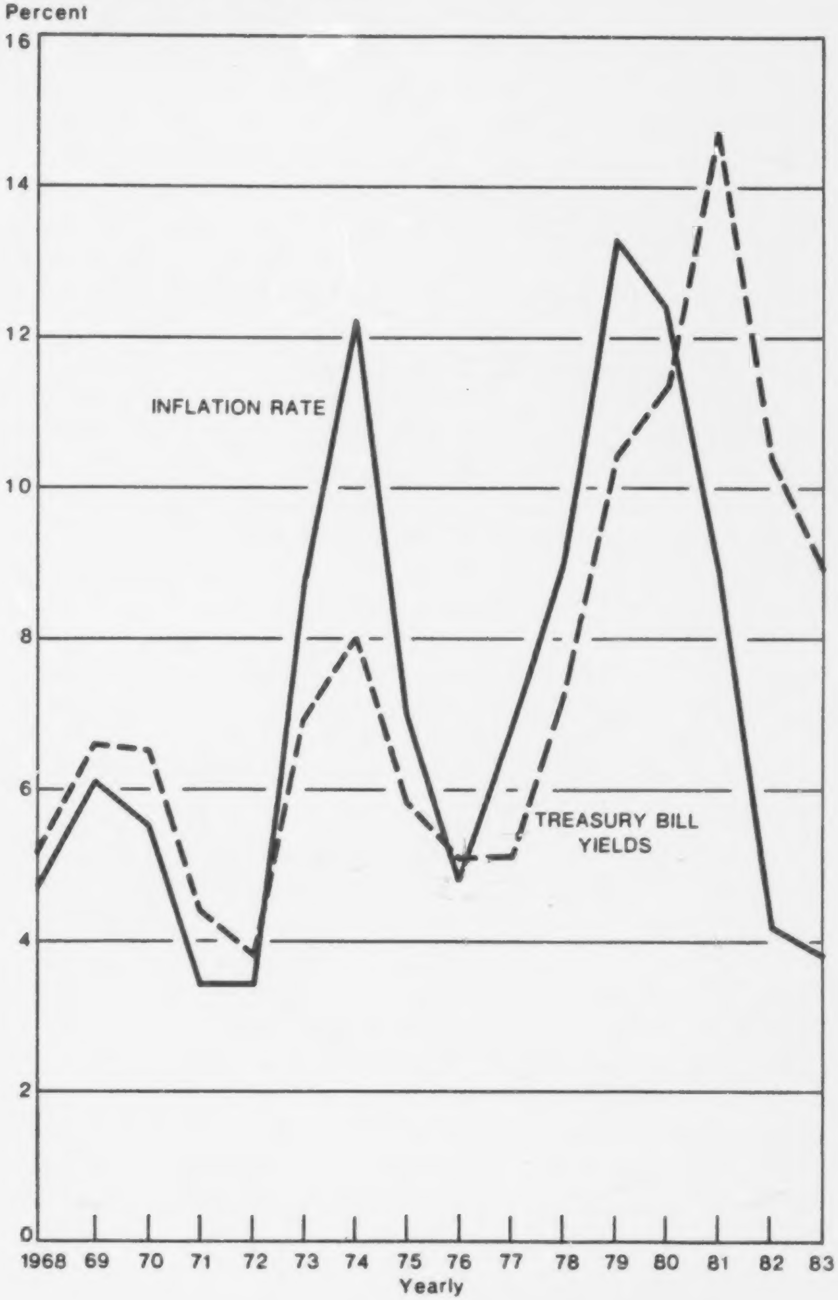
Putting the same thing somewhat differently: $1 invested in bills at the start of 1968 would have grown to $3.17 by the end of 1983; to keep exact pace with inflation and maintain the same purchasing power at the end of 1983 as at the beginning of 1968 would have required growth of $1 to $3.00.

---

[6] See Roger G. Ibbotson and Rex A. Sinquefield, *Stocks, Bonds, Bills, and Inflation: The Past and the Future* (Financial Analysts Research Foundation, 1982).

Chart 2

RATE OF RETURN ON TREASURY BILL YIELDS AND THE RATE OF INFLATION



Source: Table 3 in Chapter II and Table 7 in the Appendix to Chapter III.

Treasury bills thus provided a virtually zero *real* rate of return over the sixteen years. Compared to gold's positive 9.0 percent annual real rate of return, silver's positive 3.1 percent, and even housing's positive 0.9 percent, this does not look too rewarding. However, Treasury bills at least kept up with inflation and maintained purchasing power intact, which is more than can be said for long-term government bonds.

Stocks. In the fifties and sixties, the prevailing wisdom was that inflation was good for the stock market: that the way to combat inflation's erosion of purchasing power was to invest in stocks. The "prudent man" rule was thereupon duly reinterpreted to facilitate fiduciary acquisitions of equities in order to protect and indeed expand investor purchasing power despite rising prices. Unfortunately, the stock market disappointed many by delivering less than expected.

Tables 9 and 10 in the Appendix to this chapter show, respectively, the annual rates of return on stocks, and the amount to which $1 invested in the stock market would have grown, over the time span 1968-83. Our measure of stock market returns is calculated from a value-weighted index of all stocks on the New York Stock Exchange, which includes both capital gains (and losses) and dividends, with all dividends reinvested.

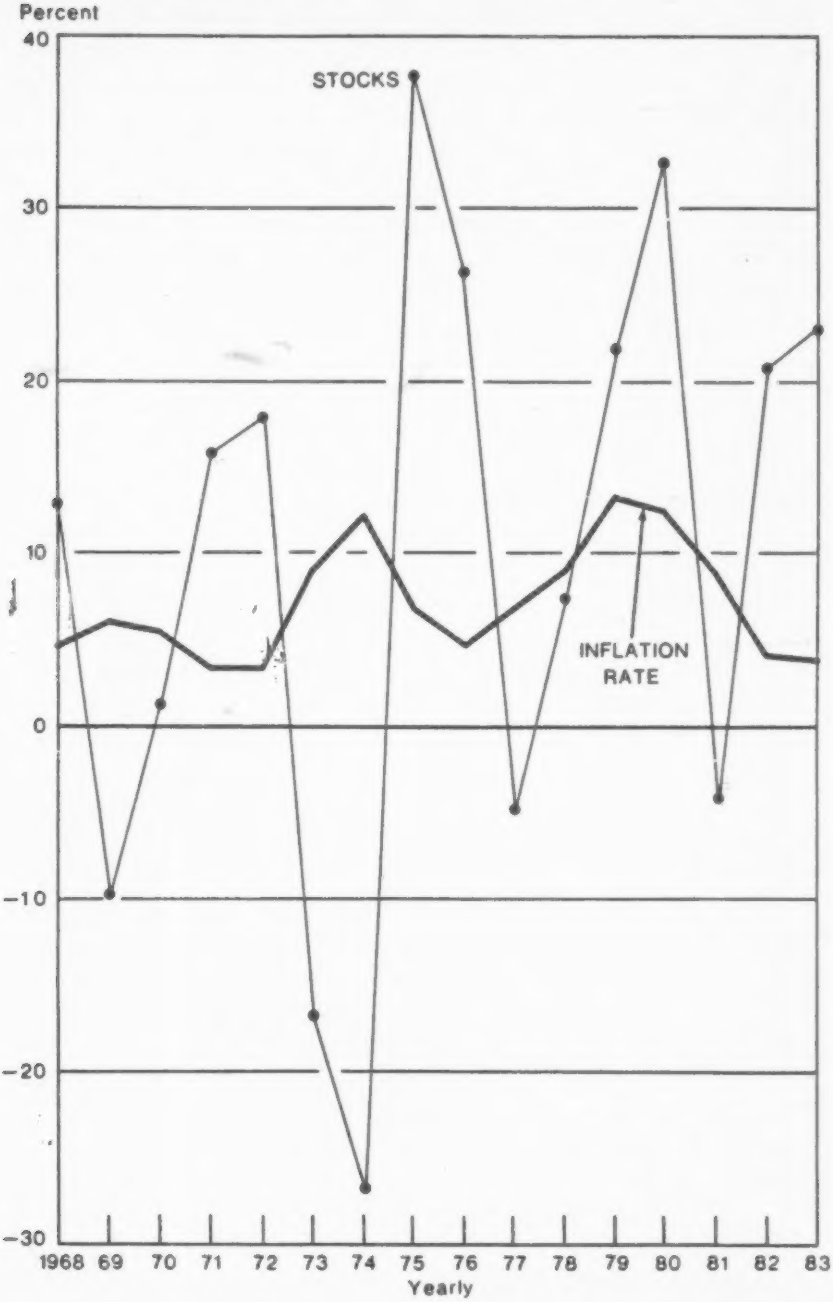
The tables show that the stock market was a poor inflation hedge for much of the period, with stock returns only 1 percent over the inflation rate after the full sixteen years. The average annual rate of return on stocks held from the start of 1968 to the end of 1983 was 8.1 percent, while the inflation rate over the same period was 7.1 percent per annum.

Nevertheless, for the holding period 1968-74 the average annual rate of return on stocks was negative (minus 2.2 percent) although the annual inflation rate over that interval was 6.3 percent. For the holding period 1968-81, stocks yielded 6.3 percent as against an inflation rate of 7.6 percent per annum. However, in the 1982 and 1983 period when inflation receeded to a 4 percent per annum, stocks yielded 21.7 percent.

Chart 3 shows the relation between the rate of return on stocks and the inflation rate on a year-to-year basis from 1968 through 1983. Inflation and stock prices moved in opposite directions more than twice as often as they moved in the same direction. By the end of the period, the Wall Street maxim "inflation is good for stock prices" had been replaced by its opposite: "inflation is bad for stock prices". (Now it is "lower interest rates" that are said to be good for stock prices.)

Chart 3

RATE OF RETURN ON STOCKS AND THE RATE OF INFLATION



Source: Table 3 in Chapter II and Table 9 in the Appendix to Chapter III.

Long-Term Government Bonds. Our measure for long-term government bonds, again following Ibbotson and Sinquefield, is a constant 20-year maturity bond portfolio. Returns include capital gains (and losses) as well as interest, with all interest reinvested. Tables 11 and 12 in the Appendix to this chapter show the rate of return and the amount to which $1 invested in long-term governments would have grown, both for the period 1968-83.

Tables 11 and 12 show that long-term government bonds have failed to keep up with inflation. Over the entire period, they yielded 5.3 percent per annum as compared with 7.1 percent inflation. To maintain constant purchasing power over the time span 1968 through 1983, $1 would have had to grow to $3.00. However, $1 invested in long-term government bonds would have grown to only $2.27; it would buy no more than what 76 cents would have bought in 1968.

The reason long-term bonds have performed so poorly is because their price is so sensitive to changes in interest rates. The longer a bond's maturity the more its price is affected, in the opposite direction, by a change in the general level of interest rates. For example, a rise in yield from 6 to 7 percent on a $10,000 face value 20-year bond, bearing a 6 percent coupon, involves a fall in price from $10,000 to $8,932. A similar rise in yield on a one-month security of similar coupon would involve a fall in price from $10,000 to only $9,990. Long-term government bonds were indeed yielding around 6 percent in the late sixties, yields that more than doubled by the early eighties with consequent erosion of capital values.

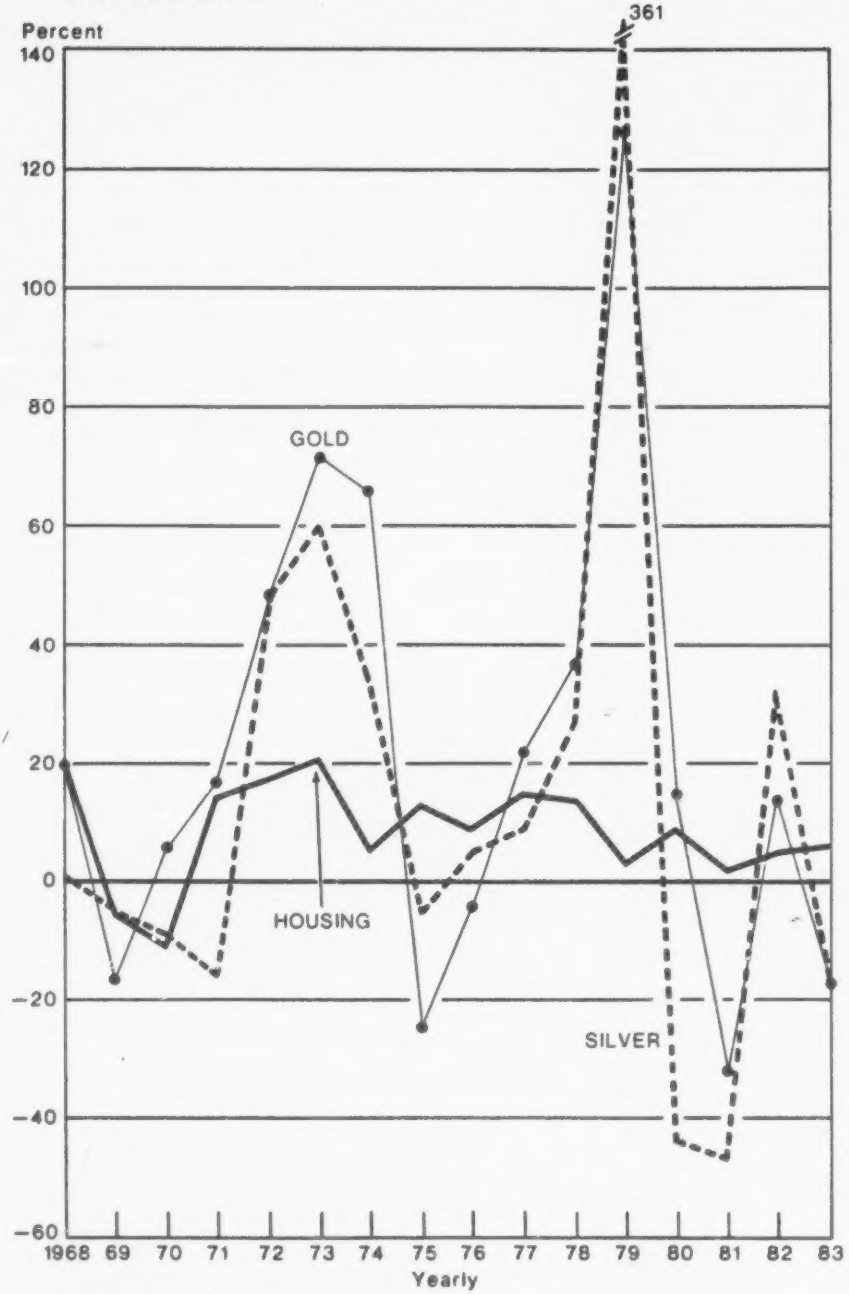
B. Relative Riskiness

It is commonplace in portfolio theory that most people are risk averse, so that higher anticipated returns are necessary to entice them to bear more risk. Since Aa bonds are poorer quality and thus riskier than Aaa bonds, for instance, Aa bonds need to carry a higher yield to induce investors to buy them; otherwise, everyone would prefer the Aaa securities. This section compares the relative riskiness of the six investments under consideration—namely, gold, silver, housing, Treasury bills, stocks, and long-term government bonds—during the period from the start of 1968 through the end of 1983.

Charts 4 and 5 show the annual rates of return on the six assets for one-year holding periods over the time span 1968-83. Chart 4 contains the three nonfinancial assets—gold, silver, and housing—while Chart 5 contains the three financial assets—stocks, long-term government bonds, and Treasury bills.

Chart 4

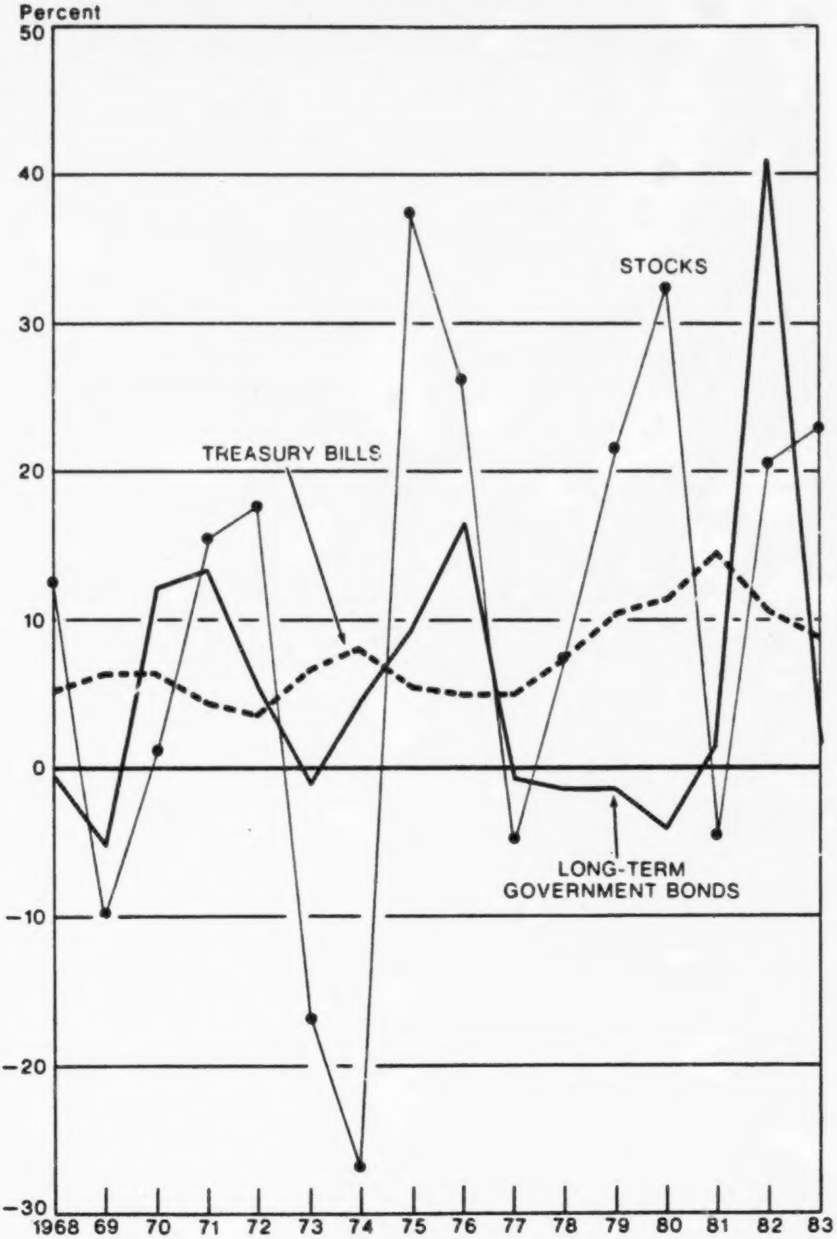
ANNUAL RATES OF RETURN ON GOLD, SILVER, AND HOUSING



Source: Table 2 in Chapter II and Tables 2 and 5 in the Appendex to Chapter III.

Chart 5

ANNUAL RATES OF RETURN ON STOCKS, LONG-TERM GOVERNMENT BONDS, AND TREASURY BILLS



Source: Tables 7, 9, and 11 in the Appendix to Chapter III.

Visual inspection of the charts indicates that gold, silver, and stocks have apparently been more volatile—and therefore riskier—than housing, long-term governments, and Treasury bills. For example, the annual rate of return on silver fluctuated between plus 361 percent (1979) and minus 47 percent (1981). Stocks fluctuated between plus 33 percent (1980) and minus 27 percent (1974). On the other hand, the rate of return on Treasury bills was much more predictable: the highest annual rate of return was 15 percent (1981) while the lowest was still positive—4 percent in 1972.

A conventional measure of risk is the standard deviation of the distribution of rates of return. As a measure of dispersion, in a normally distributed series of returns two-thirds of the returns will fall within one standard deviation on both sides of the average return, 95 percent within two standard deviations, and practically all of the returns within three standard deviations. Table 2 indicates that silver has been the most risky investment over the period, followed in order by gold, stocks long-term government bonds, housing, and Treasury bills.

TABLE 2

Standard Deviations, 1968-1983

Silver	1.36
Gold	.93
Stocks	.56
L-T Bonds	.39
Housing	.38
Bills	.03

Table 2 considers the riskiness of each asset in isolation. In the following chapter we examine them within the context of the riskiness of a total portfolio.

Appendix to Chapter III

The Underlying Data

TABLE 1

The Price of Silver Bullion, 1968-1983

	1968	1969	1970	1971	1972	1973	1974	1975
Jan.	$1.89	$1.98	$1.91	$1.60	$1.52	$2.05	$3.95	$4.10
Feb.	1.92	1.81	1.90	1.60	1.47	2.49	5.63	4.40
Mar.	2.23	1.80	1.89	1.70	1.58	2.23	5.33	4.22
Apr.	2.32	1.75	1.71	1.70	1.58	2.18	5.78	4.34
May	2.48	1.77	1.60	1.62	1.54	2.61	4.98	4.52
June	2.49	1.56	1.61	1.58	1.57	2.70	4.73	4.56
July	2.25	1.61	1.78	1.60	1.82	2.76	4.91	4.96
Aug.	2.23	1.66	1.84	1.51	1.90	2.62	4.10	4.57
Sept.	2.14	1.87	1.71	1.39	1.75	2.78	4.41	4.51
Oct.	1.92	1.97	1.80	1.31	1.83	2.86	4.91	4.16
Nov.	1.94	1.89	1.61	1.34	1.89	2.97	4.55	4.21
Dec.	1.90	1.80	1.64	1.38	2.04	3.27	4.37	4.17

TABLE 1

The Price of Silver Bullion, 1968-1983

	1976	1977	1978	1979	1980	1981	1982	1983
Jan.	$3.97	$4.47	$4.94	$ 6.75	$34.00	$13.40	$ 8.28	$13.86
Feb.	4.21	4.72	4.97	7.78	35.20	12.40	7.80	10.95
Mar.	4.08	4.79	5.41	7.43	14.25	11.90	7.06	10.40
Apr.	4.46	4.72	4.99	7.85	13.59	11.03	7.12	11.85
May	4.65	4.53	5.41	8.60	13.55	10.65	6.26	13.60
June	4.84	4.47	5.27	8.58	16.48	8.62	5.94	11.60
July	4.54	4.50	5.58	9.06	15.80	8.55	6.56	12.16
Aug.	4.15	4.45	5.51	10.48	16.15	9.19	7.97	12.04
Sept.	4.33	4.66	5.65	16.50	20.45	9.10	8.11	11.23
Oct.	4.33	4.86	6.17	16.40	18.90	9.17	10.01	8.75
Nov.	4.26	4.74	5.92	18.77	18.75	8.13	9.85	9.70
Dec.	4.38	4.78	6.07	28.00	15.65	8.25	10.90	8.95

Source: Handy and Harman. Noon price on last business day of month.

TABLE 2

Rate of Return on Silver, 1968-1983

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	0.5%															
1969	−2.4	− 5.3%														
1970	−4.7	− 7.2	− 9.2%													
1971	−7.6	−10.1	−12.4	−15.6%												
1972	1.6	1.8	4.3	11.8	48.0%											
1973	9.6	11.5	16.1	26.0	53.9	60.0%										
1974	12.7	14.9	19.4	27.9	46.9	46.3	33.7%									
1975	10.4	11.9	15.0	20.6	31.8	26.8	12.9	−4.7%								
1976	9.8	11.0	13.5	17.8	26.0	21.0	10.2	0.1	5.0%							
1977	9.7	10.8	13.0	16.6	23.0	18.5	10.0	3.0	7.1	9.3%						
1978	11.2	12.3	14.5	17.8	23.6	19.9	13.2	8.6	13.4	17.8	27.1%					
1979	25.2	27.7	31.6	37.1	45.7	45.4	43.1	45.0	61.0	85.7	142.0	361.0%				
1980	17.7	19.2	21.7	25.3	31.0	29.0	25.1	24.0	30.3	37.5	48.5	60.5	−44.1%			
1981	11.1	12.0	13.5	15.9	19.6	16.8	12.3	9.5	12.1	13.5	14.6	10.8	−45.7	−47.3%		
1982	12.4	13.3	14.9	17.1	20.7	18.2	14.3	12.1	14.7	16.4	17.9	15.7	−27.0	−16.5	32.1%	
1983	10.2	10.9	12.1	14.0	16.9	14.4	10.6	8.3	10.0	10.8	11.0	8.1	−24.8	−17.0	4.2	−17.9%

Source: Table 1.

TABLE 3

Amount to Which $1 Invested in Silver Would Grow

TO THE END OF	FROM THE BEGINNING OF 1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	$ 1.01															
1969	0.95	$ 0.95														
1970	0.87	0.86	$ 0.91													
1971	0.73	0.73	0.77	$ 0.84												
1972	1.08	1.07	1.13	1,25	$ 1.48											
1973	1.73	1.72	1.82	2.00	2.37	$ 1.60										
1974	2.31	2.30	2.43	2.67	3.17	2.14	$1.34									
1975	2.20	2.19	2.31	2.55	3.02	2.04	1.27	$0.95								
1976	2.31	2.30	2.43	2.68	3.17	2.14	1.34	1.00	$1.05							
1977	2.53	2.52	2.66	2.92	3.46	2.34	1.46	1.09	1.15	$1.09						
1978	3.21	3.20	3.37	3.72	4.40	2.97	1.86	1.39	1.46	1.39	$1.27					
1979	14.81	14.74	15.56	17.13	20.29	13.71	8.57	6.41	6.72	6.40	5.86	$4.61				
1980	8.28	8.24	8.69	9.57	11.34	7.66	4.79	3.58	3.76	3.58	3.27	2.58	$0.56			
1981	4.37	4.34	4.58	5.05	5.98	4.04	2.52	1.89	1.98	1.89	1.73	1.36	0.29	$0.53		
1982	5.77	5.74	6.06	6.67	7.90	5.34	3.34	2.49	2.62	2.49	2.28	1.79	0.39	0.70	$1.32	
1983	4.74	4.71	4.97	5.47	6.49	4.38	2.74	2.05	2.15	2.05	1.87	1.47	0.32	0.57	1.09	$0.82

Source: Tables 1 and 2.

TABLE 4

Median Sales Price of New One-Family Houses, 1968-1983

	1968	1969	1970	1971	1972	1973	1974	1975
Jan.	$23,400	$24,600	$23,600	$23,900	$24,700	$29,900	$34,200	$37,200
Feb.	23,500	26,000	24,100	24,500	26,500	29,700	34,900	37,900
Mar.	24,600	26,200	24,100	24,300	27,400	31,400	36,000	38,800
Apr.	24,500	25,900	24,200	25,800	26,700	32,800	35,700	39,200
May	25,300	26,900	25,700	25,500	27,000	32,000	35,700	39,500
June	25,400	25,200	23,800	26,100	26,800	33,100	35,100	37,900
July	24,600	26,800	22,900	25,200	27,700	34,200	36,800	38,600
Aug.	25,000	24,900	23,500	25,300	28,100	33,200	35,700	38,200
Sept.	25,000	26,000	22,600	25,400	28,000	33,200	36,200	39,700
Oct.	25,600	24,400	22,100	25,600	28,900	33,300	37,200	40,700
Nov.	24,600	24,900	23,500	25,700	29,100	34,000	37,300	41,100
Dec.	26,500	25,100	22,300	25,300	29,700	35,700	37,400	42,100

TABLE 4

Median Sales Price of New One-Family Houses, 1968-1983

	1976	1977	1978	1979	1980	1981	1982	1983
Jan.	$41,600	$45,500	$52,300	$60,300	$62,900	$67,600	$66,200	$73,500
Feb.	42,700	47,400	53,200	61,200	64,900	66,200	65,700	73,800
Mar.	43,600	46,200	53,200	60,400	62,600	66,700	67,200	72,500
Apr.	43,300	48,700	53,300	62,600	63,100	68,700	70,200	74,700
May	43,600	49,300	55,700	62,800	63,100	70,800	69,300	74,500
June	46,100	48,700	56,700	64,200	64,900	68,800	69,300	75,800
July	44,600	48,600	54,800	63,800	64,000	69,500	70,900	75,200
Aug.	44,200	49,000	56,100	64,000	63,400	72,600	70,100	76,800
Sept.	44,700	48,500	57,300	66,000	68,300	65,800	67,800	81,000
Oct.	45,300	51,400	58,300	62,300	65,600	69,600	69,700	75,900
Nov.	45,800	51,800	58,700	63,900	66,700	71,200	73,500	75,900
Dec.	45,900	52,700	59,900	61,500	67,000	68,400	71,700	75,900

Source: Department of Housing and Urban Development, Bureau of the Census, U.S. Department of Commerce.

TABLE 5

Rate of Return on New One-Family Houses, 1968-1983

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	19.4%															
1969	6.3	−5.3%														
1970	0.2	−8.3	−11.2%													
1971	3.3	−1.5	0.4	13.5%												
1972	6.0	2.9	5.8	15.4	17.4%											
1973	8.2	6.1	9.2	17.0	18.8	20.2%										
1974	7.7	5.9	8.3	13.8	13.9	12.2	4.8%									
1975	8.3	6.8	9.0	13.6	13.6	12.3	8.6	12.6%								
1976	8.4	7.1	9.0	12.8	12.7	11.5	8.7	10.8	9.0%							
1977	9.0	7.9	9.7	13.1	13.0	12.2	10.2	12.1	11.9	14.8%						
1978	9.4	8.5	10.2	13.2	13.1	12.4	10.9	12.5	12.5	14.2	13.7%					
1979	8.9	8.0	9.4	11.9	11.7	11.0	9.5	10.5	9.9	10.2	8.0	2.7%				
1980	8.9	8.0	9.3	11.6	11.4	10.7	9.4	10.2	9.7	9.9	8.3	5.8	8.9%			
1981	8.4	7.6	8.7	10.7	10.5	9.7	8.5	9.0	8.4	8.3	6.7	4.5	5.5	2.1%		
1982	8.1	7.4	8.4	10.2	9.9	9.2	8.1	8.5	7.9	7.7	6.4	4.6	5.3	3.5	4.8%	
1983	8.0	7.3	8.2	9.9	9.6	8.9	7.8	8.2	7.6	7.4	6.3	4.8	5.4	4.2	5.3	5.9%

Source: Table 4.

TABLE 6

Amount to Which $1 Invested in Housing Would Grow

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	$1.19															
1969	1.13	$0.95														
1970	1.00	0.84	$0.89													
1971	1.14	0.95	1.01	$1.13												
1972	1.34	1.12	1.18	1.33	$1.17											
1973	1.61	1.35	1.42	1.60	1.41	$1.20										
1974	1.68	1.41	1.49	1.68	1.48	1.26	$1.05									
1975	1.90	1.59	1.68	1.89	1.66	1.42	1.18	$1.13								
1976	2.07	1.73	1.83	2.06	1.81	1.55	1.29	1.23	$1.09							
1977	2.37	1.99	2.10	2.36	2.08	1.77	1.48	1.41	1.25	$1.15						
1978	2.70	2.26	2.39	2.69	2.37	2.02	1.68	1.60	1.42	1.31	$1.14					
1979	2.77	2.32	2.45	2.76	2.43	2.07	1.72	1.64	1.46	1.34	1.17	$1.03				
1980	3.02	2.53	2.67	3.00	2.65	2.26	1.88	1.79	1.59	1.46	1.27	1.12	$1.09			
1981	3.08	2.58	2.73	3.07	2.70	2.30	1.92	1.83	1.62	1.49	1.30	1.14	1.11	$1.02		
1982	3.23	2.71	2.86	3.22	2.83	2.41	2.01	1.92	1.70	1.56	1.36	1.20	1.17	1.07	$1.05	
1983	3.42	2.86	3.02	3.40	3.00	2.56	2.13	2.03	1.80	1.65	1.44	1.27	1.23	1.13	1.11	$1.06

Source: Tables 4 and 5.

TABLE 7

Rate of Return on Treasury Bills, 1968-1983

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	5.2%															
1969	5.9	6.6%														
1970	6.1	6.5	6.5%													
1971	5.7	5.8	5.5	4.4%												
1972	5.3	5.3	4.9	4.1	3.8%											
1973	5.6	5.6	5.4	5.0	5.4	6.9%										
1974	5.9	6.0	5.9	5.8	6.2	7.5	8.0%									
1975	5.9	6.0	5.9	5.8	6.1	6.9	6.9	5.8%								
1976	5.8	5.9	5.8	5.7	5.9	6.5	6.3	5.4	5.1%							
1977	5.7	5.8	5.7	5.6	5.8	6.2	6.0	5.3	5.1	5.1%						
1978	5.9	5.9	5.9	5.8	6.0	6.4	6.2	5.8	5.8	6.2	7.2%					
1979	6.2	6.3	6.3	6.3	6.5	6.9	6.9	6.7	6.9	7.6	8.8	10.4%				
1980	6.6	6.7	6.8	6.8	7.0	7.5	7.5	7.5	7.8	8.5	9.6	10.8	11.3%			
1981	7.2	7.3	7.4	7.5	7.8	8.2	8.4	8.5	8.9	9.7	10.9	12.1	13.0	14.7%		
1982	7.4	7.6	7.6	7.7	8.0	8.5	8.6	8.7	9.1	9.8	10.8	11.7	12.1	12.5	10.4%	
1983	7.5	7.6	7.7	7.8	8.1	8.5	8.6	8.7	9.1	9.7	10.4	11.1	11.3	11.3	9.6	8.9%

Source: Ibbotson and Sinquefield, *op. cit.*, and Federal Reserve Bank of New York.

TABLE 8

Amount to Which $1 Invested in Treasury Bills Would Grow

TO THE END OF	FROM THE BEGINNING OF 1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	$1.05															
1969	1.12	$1.07														
1970	1.19	1.14	$1.07													
1971	1.25	1.18	1.11	$1.04												
1972	1.29	1.23	1.15	1.08	$1.04											
1973	1.38	1.32	1.23	1.16	1.11	$1.07										
1974	1.50	1.42	1.33	1.25	1.20	1.15	$1.08									
1975	1.58	1.50	1.41	1.32	1.27	1.22	1.14	$1.06								
1976	1.66	1.58	1.48	1.39	1.33	1.28	1.20	1.11	$1.05							
1977	1.75	1.66	1.56	1.46	1.40	1.35	1.26	1.17	1.10	$1.05						
1978	1.87	1.78	1.67	1.57	1.50	1.45	1.35	1.25	1.18	1.13	$1.07					
1979	2.07	1.97	1.84	1.73	1.66	1.60	1.49	1.38	1.31	1.24	1.18	$1.10				
1980	2.30	2.19	2.05	1.93	1.85	1.78	1.66	1.54	1.45	1.38	1.32	1.23	$1.11			
1981	2.64	2.51	2.35	2.21	2.12	2.04	1.91	1.77	1.67	1.59	1.51	1.41	1.28	$1.15		
1982	2.91	2.77	2.60	2.44	2.34	2.25	2.10	1.95	1.84	1.75	1.67	1.56	1.41	1.27	$1.10	
1983	3.17	3.02	2.83	2.66	2.54	2.45	2.29	2.12	2.01	1.91	1.82	1.69	1.54	1.38	1.20	$1.09

Source: Table 7.

TABLE 9

Rate of Return on Stocks, 1968-1983

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	12.8%															
1969	0.9	−9.8%														
1970	1.0	−4.4	1.3%													
1971	4.5	1.9	8.3	15.8%												
1972	7.1	5.7	11.4	16.8	17.8%											
1973	2.6	0.7	3.5	4.3	− 1.1	−16.9%										
1974	−2.2	−4.5	−3.4	−4.6	−10.5	−22.0	−26.8%									
1975	2.1	0.6	2.5	2.7	− 0.3	−5.7	0.4	37.7%								
1976	4.5	3.5	5.6	6.3	4.5	1.4	8.4	31.9	26.3%							
1977	3.5	2.6	4.2	4.6	2.9	0.1	4.9	18.3	9.6	−4.8%						
1978	3.9	3.0	4.6	5.0	3.5	1.3	5.4	15.5	8.9	1.1	7.4%					
1979	5.3	4.6	6.2	6.7	5.7	4.0	8.0	16.7	12.0	7.6	14.4	21.9%				
1980	7.2	6.7	8.4	9.1	8.4	7.2	11.2	19.3	15.9	13.4	20.2	27.2	32.7%			
1981	6.3	5.8	7.2	7.8	7.0	5.9	9.2	15.6	12.3	9.6	13.6	15.7	12.7	−4.2%		
1982	7.2	6.8	8.2	8.8	8.2	7.3	10.4	16.2	13.4	11.4	15.0	17.0	15.3	7.5	20.7%	
1983	8.1	7.8	9.2	9.8	9.3	8.6	11.6	16.9	14.5	13.0	16.2	18.1	17.1	12.4	21.7	23.0 %

Source: Center for Research in Securities Prices, Graduate School of Business, University of Chicago.

TABLE 10

Amount to Which $1 Invested in Stocks Would Grow

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	$1.13															
1969	1.02	$0.90														
1970	1.03	0.91	$1.01													
1971	1.19	1.06	1.17	$1.16												
1972	1.41	1.25	1.38	1.36	$1.18											
1973	1.17	1.04	1.15	1.13	0.98	$0.83										
1974	0.85	0.76	0.84	0.83	0.72	0.61	$0.73									
1975	1.18	1.04	1.16	1.14	0.99	0.84	1.01	$1.38								
1976	1.49	1.32	1.46	1.44	1.25	1.06	1.27	1.74	$1.26							
1977	1.42	1.25	1.39	1.37	1.19	1.01	1.21	1.66	1.20	$0.95						
1978	1.52	1.35	1.49	1.47	1.27	1.08	1.30	1.78	1.29	1.02	$1.07					
1979	1.85	1.64	1.82	1.80	1.55	1.32	1.59	2.17	1.57	1.25	1.31	$1.22				
1980	2.46	2.18	2.42	2.39	2.06	1.75	2.11	2.88	2.09	1.65	1.74	1.62	$1.33			
1981	2.36	2.09	2.31	2.28	1.97	1.68	2.02	2.75	2.00	1.58	1.66	1.55	1.27	$0.96		
1982	2.84	2.52	2.79	2.76	2.38	2.02	2.43	3.33	2.41	1.91	2.01	1.87	1.53	1.16	$1.21	
1983	3.49	3.09	3.43	3.39	2.92	2.48	2.99	4.08	2.96	2.35	2.47	2.30	1.88	1.42	1.48	$1.23

Source: Table 9.

TABLE 11

Rate of Return on Long-Term Government Bonds, 1968-1983

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	−0.3%															
1969	−2.7	−5.1%														
1970	2.0	3.2	12.1%													
1971	4.7	6.4	12.7	13.2%												
1972	4.9	6.2	10.3	9.4	5.7%											
1973	3.9	4.7	7.3	5.8	2.2	−1.1%										
1974	3.9	4.7	6.7	5.4	2.9	1.6	4.4%									
1975	4.6	5.3	7.1	6.2	4.5	4.1	6.8	9.2%								
1976	5.9	6.7	8.5	7.9	6.8	7.1	10.0	12.9	16.8%							
1977	5.2	5.8	7.3	6.6	5.5	5.5	7.2	8.2	7.7	−0.7%						
1978	4.6	5.1	6.3	5.6	4.6	4.4	5.5	5.8	4.7	−0.9	−1.2%					
1979	4.1	4.5	5.5	4.8	3.8	3.6	4.3	4.3	3.2	−1.0	−1.2	−1.2%				
1980	3.5	3.8	4.6	3.9	2.9	2.6	3.1	2.9	1.7	−1.8	−2.1	−2.6	−3.9%			
1981	3.4	3.6	4.4	3.7	2.8	2.5	3.0	2.8	1.7	−1.0	−1.1	−1.1	−1.1	1.9%		
1982	5.5	5.9	6.8	6.4	5.8	5.8	6.6	6.9	6.6	5.0	6.1	8.1	11.3	19.9	41.0%	
1983	5.3	5.6	6.5	6.0	5.5	5.4	6.1	6.3	6.0	4.5	5.4	6.7	8.8	13.4	19.7	1.6%

Source: Ibbotson and Sinquefield, *op. cit.*, and Salomon Brothers.

TABLE 12

Amount to Which $1 Invested in Long-Term Bonds Would Grow

	FROM THE BEGINNING OF															
TO THE END OF	1968	'69	'70	'71	'72	'73	'74	'75	'76	'77	'78	'79	'80	'81	'82	'83
1968	$1.00															
1969	0.95	$0.95														
1970	1.06	1.06	$1.12													
1971	1.20	1.20	1.27	$1.13												
1972	1.27	1.27	1.34	1.20	$1.06											
1973	1.26	1.26	1.33	1.18	1.05	$0.99										
1974	1.31	1.31	1.38	1.23	1.09	1.03	$1.04									
1975	1.43	1.43	1.51	1.35	1.19	1.13	1.14	$1.09								
1976	1.67	1.68	1.77	1.57	1.39	1.32	1.33	1.28	$1.17							
1977	1.66	1.66	1.75	1.56	1.38	1.31	1.32	1.27	1.16	$0.99						
1978	1.64	1.64	1.73	1.55	1.37	1.29	1.31	1.25	1.15	0.98	$0.99					
1979	1.62	1.62	1.71	1.53	1.35	1.28	1.29	1.24	1.13	0.97	0.98	$0.99				
1980	1.56	1.56	1.64	1.47	1.30	1.23	1.24	1.19	1.09	0.93	0.94	0.95	$0.96			
1981	1.59	1.59	1.68	1.49	1.32	1.25	1.26	1.21	1.11	0.95	0.96	0.97	0.98	$1.02		
1982	2.24	2.24	2.36	2.11	1.86	1.76	1.78	1.71	1.56	1.34	1.35	1.36	1.38	1.44	$1.41	
1983	2.27	2.28	2.40	2.14	1.89	1.79	1.81	1.73	1.59	1.36	1.37	1.39	1.40	1.46	1.43	$1.02

Source: Table 11.

IV. Looking Forward: The Role of Gold In a Modern Investment Portfolio

A. Gold as a Hedge Against Inflation

From 1950 through 1983, the cost of living in the United States, as measured by the Consumer Price Index, increased in every year but one (1955). The annual rate of inflation over the entire 33-year period averaged 4½ percent. Although the inflation rate averaged 2.2 percent a year during the fifties and sixties, it tripled to over 7 percent a year in the seventies. Inflation reached double-digit figures three times in this time span: in 1974 (12.2 percent), 1979 (13.3 percent), and 1980 (12.4 percent).

It is apparent that the pace of inflation has accelerated over the past three decades. Although it is impossible to know what will happen over the remaining years of the twentieth century, not to mention thereafter, prudence dictates that a hedge against inflation be built into modern investment portfolios. Inflation at the rate of 5 percent a year doubles the price level every 14 years; at the rate of 7 percent annually it doubles the price level every 10 years. In light of recent experience, any portfolio management that ignores the potential wealth-eroding effects of continued inflation should be held grossly negligent.

A generation ago, the stock market was widely viewed as a good hedge against inflation. However, as indicated in Chapter III, the stock market has yielded disappointing results. In *real* terms, the Dow-Jones Industrial Average peaked in the mid-sixties and has yet to fully recover. Over the 16-year period from the start of 1968 to the end of 1983, the *real* annual rates of return on the six assets considered in the previous chapters were as shown in Table 1.

Gold has fared much better than the other listed assets in the inflationary environment of 1968-83. Whether or not it will continue to do as well in the future no one can say with certainty, but of the six assets listed it has been the best inflation hedge over the past decade and a half.

The value of gold as an inflation hedge over the period 1968-83 is corroborated by the coefficients of correlation between the rate of inflation and rates of return on these assets over the same time period. Table 2 shows such correlation coefficients calculated in two different ways. The first column contains correlation coefficients calculated on the basis of monthly data. The second column contains similar coefficients but calculated on the basis of annual data.

TABLE 1

Real Rates of Return, 1968-1983

Asset	Nominal Rate of Return	Minus	Rate of Inflation	Equals	Real Rate of Return
Gold	16.1%		7.1%		9.0%
Silver	10.2		7.1		3.1
Stocks	8.1		7.1		1.0
Housing	8.0		7.1		.9
Bills	7.5		7.1		.4
L-T Bonds	5.3		7.1		−1.8

TABLE 2

Correlation Coefficients, 1968-1983

Correlation between Inflation and:	Based on Monthly Data	Based on Annual Data
Bills	.43	.53
Gold	.12	.52
Silver	.07	.44
Housing	−.04	−.12
L-T Bonds	−.12	−.47
Stocks	−.19	−.22

A positive correlation coefficient means that two variables—such as the rate of inflation and the rate of return on Treasury bills—move in the same direction, while a negative coefficient means that as one variable goes up the other goes down. On the basis of correlation coefficients, the best inflation hedges are bills, gold, and silver. Housing, bonds, and stocks have negative coefficients, whether calculated with monthly or actual data.[7]

---

[7] For those interested, a full correlation matrix containing all the variables involved looks like the following (annual data, 1968-1983):

	Bills	Gold	Silver	Housing	Bonds	Stocks	CPI
Bills	1.00						
Gold	− .05	1.00					
Silver	.10	.82	1.00				
Housing	− .39	.22	− .04	1.00			
Bonds	− .01	− .18	− .09	− .13	1.00		
Stocks	− .02	− .22	.06	.15	.28	1.00	
CPI	.53	.52	.44	− .12	− .47	− .22	1.00

Additional confirmation of gold as an inflation hedge comes from the widely-quoted Salomon Brothers annual survey of annual rates of return on a wide variety of assets. Table 3 contains the results of the Salomon Brothers survey for periods of 15, 10, and 5 years ending in mid-1983. The numbers differ somewhat from the numbers we have been using because of differences in dating—the Salomon survey uses mid-year figures rather than end-of-year data—and differences in measurement methods.

As Table 3 indicates, over the 15-year time span gold ranked fourth, over the 10-year time span fifth, and over the 5-year time span third among the fourteen investments surveyed. Rare coins and stamps, the only two consumer investment-type assets yielding higher returns than gold over the full 15 years, both require specialized knowledge and expertise not needed for investment in gold.

B. Diversification with Gold Can Reduce Risk

Another way of viewing gold as an investment is to examine the riskiness of a total portfolio with and without the inclusion of gold. At first glance, it would appear that adding gold to a portfolio would make the portfolio more risky, because gold itself is among the more risky investments. We saw in Chapter III, for example, that gold, with a 93 percent standard deviation of its rates of return over the years 1968-83, was second only to silver in variability.

Nevertheless, under many circumstances the addition of gold can *reduce* total portfolio risk because the rate of return on gold fluctuates *opposite* to the rate of return on stocks. Based on annual data from 1968 through 1983, the correlation coefficient between gold and the stock market is $-.22$: when inflation accelerates, the price of gold typically goes up while the stock market typically goes down; when inflation subsides, the price of gold typically falls while the stock market typically rises.

The beta of an asset is a measure of its sensitivity to movements in the market as a whole. A beta coefficient greater than 1 means that an asset's returns are more volatile than rates of return for the market as a whole, thereby increasing the overall riskiness of a portfolio.[8] A beta coefficient less than 1 means that an asset's returns are less volatile than the market as a whole, thereby decreasing overall portfolio risk. A zero beta indicates a riskless asset whose returns do not vary at all with the market as a whole. Treasury bills, for example, with a beta typically close to zero, are widely considered the riskless asset *par excellence*.

---

[8] Volatility is that part of total variablity of an asset due to co-movements with the market portfolio. It is measured by an asset's beta and indicates the amount of risk that an asset adds to the overall market portfolio.

TABLE 3

Salomon Brothers Survey: Annual Rates of Return

	1968-1983		1973-1983		1978-1983	
	Rate of Return	Rank	Rate of Return	Rank	Rate of Return	Rank
Oil	20.4%	1	25.4%	2	16.2%	4
U.S. Coins	17.9	2	25.7	1	13.2	6
U.S. Stamps	16.8	3	19.2	3	21.8	1
Gold	16.6	4	15.5	5	17.5	3
Chinese Ceramics	14.2	5	4.0	14	13.1	7
Silver	12.6	6	17.3	4	19.7	2
Diamonds	10.1	7	10.3	7	5.4	13
Farmland	10.0	8	11.7	6	7.0	12
Treasury Bills	8.8	9	10.1	8	12.8	8
Housing	8.6	10	9.2	9	7.4	10
Old Masters	7.8	11	8.4	10	4.1	14
Cons. Price Index	7.3	12	8.5	11	9.1	9
Bonds	6.4	13	6.6	13	7.2	11
Stocks	5.7	14	7.5	12	14.8	5
Foreign Exchange	3.1	15	1.4	15	−2.8	15

Source: Salomon Brothers Inc. All returns are for the period ended June 1, 1983.

Table 4 shows betas, calculated on the basis of monthly data, for bills, housing, gold, bonds, and silver, each relative to the stock market as a whole (with the stock market represented by a value-weighted index of all stocks on the New York stock exchange). The table contains betas calculated using the full 16-year period, 1968-1983, and also for two five-year subperiods, 1968-72 and 1973-77, and the remaining six-year subperiod 1978-83.

TABLE 4

Betas Relative to the Stock Market

	1968-1983	1968-72	1973-77	1978-83
Bills	−.004	−.004	−.01	− .01
Housing	.05	−.06	−.01	.18
Gold	.16	.04	−.15	.58
L-T Bonds	.25	.22	.13	.40
Silver	.48	.37	.03	1.06

For the full 16-year time span, the replacement of part of an all-stock portfolio with gold would have reduced total risk: the beta of .16 for gold means that on average every time the stock market went down 10 percent gold fell only 1.6 percent, while every time the market went up 10 percent gold rose only 1.6 percent. Of course, the replacement of part of an all-stock portfolio with Treasury bills and housing would have reduced risk exposure even more, since their betas of close to zero indicate that their returns are virtually independent of what happens in the stock market.

The fact of the matter is that most people do not hold all of their investments in the form of stocks. They have some stocks, but they also own other investment-type assets: their homes, for instance, and if account is taken of their pension funds, they own a fair amount of long-term bonds as well. For this reason, it is unrealistic for individuals to calculate betas only with reference to the stock market when considering whether including gold in a consumer investment portfolio would increase or decrease total portfolio risk.

Table 5 shows betas again, calculated once more on the basis of monthly data. For Table 5, however, they have not been calculated relative to the stock market as a whole. Instead, the "market as a whole" has been redefined as a package consisting of one-third stocks, one-third long-term bonds, and one-third housing, a bundle of assets that is more representative of a typical consumer investment portfolio that 100 percent investment in the stock market. The table again contains betas for the full 16-year period, 1968-1983, and the three sub-periods presented in Table 4.

TABLE 5

Betas Relative to Stocks, Bonds, and Housing (⅓ Each)

	1968-1983	1968-72	1973-77	1978-83
Bills	−.01	−.01	−.01	− .01
Gold	.41	.19	−.22	.80
Silver	.69	.20	−.06	1.25

The betas in Table 5 are larger than in Table 4. They measure the sensitivity of an asset to movements in a total package consisting of equally proportioned dollar amounts of stocks, bonds, and housing. A beta coefficient greater (less) than 1 means than an asset's returns are more (less) volatile than rates of return for the total stocks-bonds-housing package, thereby increasing (decreasing) overall portfolio risk.

For the full 16-year time span, replacement of part of an equally proportioned stocks-bonds-housing portfolio with gold would again have reduced total risk: the beta of .41 for gold means that on average every time the stocks-bonds-housing package went up (or down) 10 percent, gold rose (or fell) only 4.1 percent.

C. How Much Gold Should Be in a Modern Consumer Portfolio?

We have seen that over the period 1968-1983 gold has been a high-yielding investment, with a nominal rate of return of 16.1 percent per annum and a real rate of return of 9.0 percent. We have also seen that over these 16 years gold has provided a good hedge against inflation. In addition, although it is risky *per se*, when added to a conventional portfolio it has provided diversification that *reduces* rather than increases total portfolio risk. Under the circumstances, it appears obvious that the historical record justifies a role for gold in a modern portfolio. The only question is: how much of a role? How much gold should there be in a modern consumer portfolio?

The answer to such a question has to be qualified. We have been interpreting historical relationships and the past is not always a good guide to the future. Will world political conditions stabilize or continue unsettled? Will inflation be as serious a problem in the future as it has been in the past decade and a half? Will real interest rates remain at the relatively high levels of the early eighties or return to the relatively low levels of the late seventies? These and related questions are paramount when it comes to evaluating the place of gold in a consumer investment portfolio.

Everyone will answer such questions their own way. Someone who anticipates increased worldwide political tensions, worse inflation, and low real interest rates will naturally see gold playing a larger role in his or her portfolio than someone who expects detente, less inflation, and high real interest rates.

Not knowing the future any better than anyone else, in place of forecasting we have made an attempt to quantify the portfolio implications of the past. Assuming that the remaining years of the century resemble the period 1968-83, we have constructed Tables 6 and 7 based on the past rates of return and standard deviations of the assets we have been considering in this monograph.

Table 6 shows rates of return and corresponding risk levels, as measured by the standard deviation of the total portfolio, of alternative investments in (a) the stock market, (b) gold, and (c) Treasury bills. For example, a portfolio invested 100 percent in the stock market would yield an 8.12 percent annual rate of return with an accompanying risk level, as measured by the standard deviation, of 55.76 percent (see top left-hand corner box).

Moving to the right along the top row of Table 6 shows the implications of replacing stocks with gold. Investing 75 percent of a portfolio in gold instead of stocks, for instance, raises the annual rate of return from 8.12 percent to 14.08 percent, and simultaneously

TABLE 6

Portfolio Alternatives Among Stocks, Bills, and Gold

		25% Gold	50% Gold	75% Gold	100% Gold
100% Stocks	8.12%	10.11%	12.09%	14.08%	16.06%
	55.76	49.72	56.33	72.19	92.67
25% Bills	7.96	9.95	11.93	13.92	
	41.78	37.85	49.57	69.43	
50% Bills	7.80	9.79	11.77		
	27.82	28.02	46.20		
75% Bills	7.64	9.63			
	13.96	23.03			
100% Bills	7.48				
	2.91				

increases the level of risk from 55.76 to 72.19 percent. A portfolio that is invested 100 percent in gold would yield 16.06 percent and carry a risk equal to 92.67 percent (see top right-hand corner box).

Moving *down* along the left-hand column of Table 6 shows the implications of replacing stocks not with gold but with Treasury bills. Indeed, moving down along any column always shows the implications of adding more bills. Investing 75 percent of a portfolio in Treasury bills in place of stocks, for instance, lowers the annual rate of return from 8.12 percent to 7.64 percent, and also lowers risk from 55.76 to 13.96 percent. A portfolio invested 100 percent in Treasury bills would yield 7.48 percent and carry a level of risk equal to 2.9 percent (see bottom lefthand corner box).

The boxes in the remainder of the table show the implications of allocating a portfolio among some of each of the three alternatives. Moving to the right shows the effects of adding gold, moving down the effects of adding Treasury bills. For example, as we have seen, investing 75 percent of a portfolio in gold, with 25 percent remaining in the stock market, would yield 14.08 percent at a risk level of 72.19 percent. Taking 25 percent of that portfolio out of stocks and moving it into Treasury bills would result in a portfolio composed of 75 percent gold and 25 percent bills; according to Table 6, such a portfolio would yield 13.92 percent annual rate of return at a risk level of 69.43 percent.

Alternatives like those presented in Table 6 cannot by themselves determine an optimum portfolio, because people differ in the amount of risk they want to bear. Aggressive investors, willing to bear more risk in hopes of obtaining a higher yield, will buy more gold; defensive investors, who prefer a more certain yield even though it may be lower, will buy more Treasury bills. For any given rate of return, however, less risk is always preferable to more; and for any given level of risk, a higher return is always preferable to a lower one.

According to Table 6, moving to the right along the top row and giving up some stocks for gold will increase yield *and* reduce risk up to the combination of 25 percent gold-75 percent stocks. At that combination one can expect a 10.11 percent annual rate of return at a risk level of 49.72 percent.

However, Table 6 is unrealistic for consumer portfolios because it is built on the assumption of an all-stock portfolio. As we have observed earlier in this chapter in conjunction with betas, many people do not hold all of their investments in the form of stocks; they also own their own homes and long-term bonds (in their pension funds).

TABLE 7

Portfolio Alternatives Among Stocks-Bonds-Housing, Bills, And Gold

		5% Gold	10% Gold	15% Gold	20% Gold	25% Gold	30% Gold	35% Gold	40% Gold
100% Stocks-Bonds-Housing	7.12	7.57	8.01	8.46	8.91	9.35	9.80	10.25	10.70
	29.85	29.33	29.56	30.51	32.11	34.29	36.93	39.95	43.26
5% Bills	7.14	7.58	8.03	8.48	8.92	9.37	9.82	10.27	10.71
	28.35	27.85	28.13	29.17	30.89	33.18	35.94	39.06	42.47
10% Bills	7.15	7.60	8.05	8.50	8.94	9.39	9.84	10.28	10.73
	26.85	26.37	26.71	27.85	29.69	32.10	34.98	38.22	41.73
15% Bills	7.17	7.62	8.07	8.51	8.96	9.41	9.86	10.30	10.75
	25.35	24.89	25.31	26.55	28.51	31.06	34.07	37.41	41.02
20% Bills	7.19	7.64	8.08	8.53	8.98	9.43	9.87	10.32	10.77
	23.86	23.42	23.91	25.27	27.37	30.06	33.19	36.66	40.36
25% Bills	7.21	7.66	8.10	8.55	9.00	9.44	9.89	10.34	10.79
	22.36	21.95	22.53	24.02	26.27	29.10	32.37	35.94	39.75
30% Bills	7.23	7.67	8.12	8.57	9.02	9.46	9.91	10.36	10.80
	20.87	20.49	21.17	22.80	25.21	28.20	31.59	35.28	39.18
35% Bills	7.25	7.69	8.14	8.59	9.03	9.48	9.93	10.38	10.82
	19.38	19.04	19.83	21.62	24.20	27.34	30.87	34.67	38.67
40% Bills	7.26	7.71	8.16	8.60	9.05	9.50	9.95	10.39	10.84
	17.89	17.59	18.51	20.48	23.24	26.54	30.21	34.12	38.21

Table 7 has been constructed on the same principles as Table 6 except that the all-stock alternative has been replaced by a package consisting of one-third stocks, one-third long-term government bonds, and one-third housing (in the same way as when we calculated betas—see Table 5 as compared with Table 4). The risk levels are generally lower in Table 7 than in Table 6, and the returns as well, because the market return used to construct Table 7 has more diversification built into it than the market return in Table 6.

Table 7 indicates that moving to the right along the top row and replacing 5 percent of the stocks-bonds-housing package with gold will increase yield *and* lower risk. Moving to 10 percent of the portfolio in gold will increase yield further, and increase risk as well, but overall portfolio risk will still be less than that of a 100 percent stocks-bonds-housing portfolio.

On average, a benchmark range of 5 to 10 percent of one's total investment portfolio in gold appears warranted and prudent in light of the historical record. Within that range, the proportion that is appropriate for any particular individual depends on such factors as how one views the future—in terms of the outlook for world tensions, inflation, and the level of real interest rates—and how much risk one feels comfortable bearing.

Those who anticipate increased world tensions, rising inflation, and low real interest rates, and who feel comfortable bearing risk, will want to go to the 10 percent level if not beyond. While those who expect an easing of world tensions, less inflation, high real interest rates, and who dislike risk, may well find 5 percent too much. For most investors, however, the evidence indicates that somewhere between 5 and 10 percent of one's portfolio in gold is appropriate.

Appendix to Chapter IV:

Transactions Costs Associated with Different Forms of Gold

This appendix discusses the transactions costs associated with investing in gold. The cheapest way to purchase gold, in terms of transactions costs per ounce, is 100-ounce bars—both because the volume of ounces purchased lowers the cost per ounce and because of the enhanced liquidity of the market for bars of this size. Small investors who do not want to purchase gold in a 100-ounce size have to pay higher transactions costs per ounce. Below, we give examples of the typical costs associated with a round-trip transaction of purchasing and then selling gold in 100-ounce bars, 1-ounce bars, gold coins and accumulation plans.

1) ***100-ounce gold bars.*** 100-ounce gold bars are the cheapest way to purchase gold per ounce. The cost of buying or selling these bars is 1% of the spot price plus $130. Assuming a spot price of $400 an ounce at the London fix, the purchase price per ounce would be $405.30 and the selling price per ounce would be $394.70, giving a round-trip cost of $10.60 per ounce:

	Purchase Price	Sales Price
Spot Price/ounce	$400.00	$400.00
× 100 ounces	100	100
Spot Price per bar	$40,000	$40,000
× 1%	+ 400	− 400
$130	+ 130	− 130
Price per 100 ounces	$40,530	$39,470
÷ 100	÷ 100	÷ 100
Cost per ounce	$405.30	$394.70

Cost of a round-trip transaction = $405.30 − 394.70 = $10.60

2) ***Accumulation Plans.*** In an accumulation plan the investor agrees to make regular purchases of a minimum dollar amount of gold for a specific time interval—say a minimum of $2000 per

month for six months. The broker pools the small order together with all the other accumulation plans to form large purchases (or sales) that are executed in the 100-ounce bar market to reduce the transactions cost per ounce. The benefit of the accumulation plan is that the investor is able to buy small fixed-dollar amounts of gold with relatively low transactions costs. The purchase price in the accumulation plan is $1 over the spot price plus a 2% commission. The sales price is $1 under spot minus a 2% commission. If the spot price was $400 the purchase price per ounce would be $409.02 while the sales price would be $391.02, giving an $18 total cost:

	Purchase	Sale
Spot price	$400.00	$400.00
$1 differential	+ 1.00	− 1.00
2% commission	+ 8.02	− 7.98
	$409.02	$391.02

Cost of a round-trip transaction = $409.02 − 391.02 = $18

3) ***Small Gold Bars.*** The most liquid size of small bars (bars less than 100 ounces) is the 1-ounce bar. The purchase price for a minimum purchase of 10 one-ounce bars, is 1.25% over spot plus a 2% commission. The sales price is 2% under spot. Given a $400 spot price for gold the purchase and sale price are $413.10 and $392 respectively, giving a round-trip cost of $21.10:

	Purchase Price	Sales Price
Spot	$400.00	$400.00
Fabrication cost 1.25%	+ 5.00	
2% Commission	+ 8.10	− 8.00
	$413.10	$392.00

Cost of a round-trip transaction = $413.10 − 392.00 = $21.10

4) ***Gold Coins.*** The two major gold coins are the South African Krugerrand and the Canadian Maple Leaf. In the United States these coins (and also probably the Mexican Peso and Austrian Korona) trade at the same price. However, in the international market (e.g.

Europe) the Krugerrand tends to be more widely recognized because of substantial marketing and greater volume outstanding. The result is that the Krugerrand may tend to maintain its premium slightly better than some of the other coins. The purchase price of one-ounce gold coins (for a minimum of five coins) is 3.5% over spot plus a 2% commission. The sales price would be 3.5% (to 3%) over spot minus $5 (bid/ask spread) minus a 2% commission. Assuming a $400 spot price the purchase and sales price would be $422.28 and $400.82 respectively, giving a $21.46 round-trip cost:

Purchase Price		Sales Price	
Spot	$400.00	Spot	$400.00
Premium 3.5%	+ 14.00	Premium	+ 14.00
2% Commission	+ 8.28	Bid/ask spread	− 5.00
	$422.28	2% Commission	− 8.18
			$400.82

Cost of a round-trip transaction = $422.28 − 400.82 = $21.46

These examples give the costs for a round-trip transaction assuming that the investor does not take physical delivery of the gold. If the investor takes physical delivery, he must pay local sales tax. Alternatively, he can avoid sales tax and leave the gold with a depository which is usually insured. The storage costs are relatively small.

Additionally, per ounce commission on gold purchases decrease as the volume of ounces increases. The pricing formulas given above for the cost of investing in gold will also change if the per ounce price of gold changes substantially.

There are also several ways to indirectly invest in gold. For example, an investor may purchase stock in gold mining companies, futures contracts in gold, and options contracts in gold or in gold futures. However, a full description of these methods is beyond the scope of this monograph.

V. Summary

Under the traditional domestic and international gold standard, each nation used to set an official price for gold and then stand ready to buy or sell unlimited amounts at that price. The market price thus became identical with the official price.

The United States went off the domestic gold standard in the early 1930s but remained on the gold standard internationally, with gold officially priced at $35 an ounce. The U.S. Treasury did business at that price only with foreign governments and official international monetary institutions. Gold coin and bullion were not allowed in private hands domestically.

In 1961, faced with a heavy private demand for gold in London and Zurich, eight of the world's leading financial nations established a "gold pool" to keep the market price from rising above the $35-an-ounce official price. The gold pool broke down in March of 1968 and was replaced by a "two-tier" gold price system, with all official transactions taking place at $35 an ounce but private transactions able to take place at a market price that was free to vary with supply and demand.

The two-tier system also broke down in 1973 and with it what was left of the international gold standard. Since then, gold has been free to find its own price, unencumbered by governmental interference or any "official" price. And since December 31, 1974, Americans have once again been allowed to own gold.

It is not likely that the United States will return to a gold standard in any form in the foreseeable future. Thus there will be no "official" price of gold but only a free market price determined by supply and demand.

With respect to supply and demand, the overwhelming element on the supply side is not the amount of current output, which is usually only about 1 to 2 percent of the total stock in existence, but uncertainty regarding how much holders of the existing stockpile might decide to sell. On the demand side, a variable portion of the demand for gold is motivated by psychological considerations—in particular, by periodic fears regarding an uncertain social and economic future. Traditionally, gold has been considered a hedge against inflation, with demand expanding when inflation accelerates and contracting when inflation subsides.

From early 1968, when the central banks abandoned fixing the price of gold through their gold pool, to the end of 1983, six major trends are discernible in the price of gold: (1) From early 1968 to the end of 1974, the market price rose from $35 an ounce to $197; (2) From January 1975 to August 1976, the price fell from almost $200 an ounce to just over $100; (3) From August 1976 to near the end of January 1980, the price skyrocketed from $104 to an all-time high of $875 an ounce; (4) From January 1980 to June 1982, the price retreated irregularly from its $875 an ounce peak to slightly below $300; (5) From June 1982 to the close of the year, the price of gold rose from $296 an ounce to $450 an ounce; (6) From January 1983 to December 1983, the price of gold see-sawed on a downward trend from a peak of about $500 in January to about $380 an ounce at year end.

Overall, from the beginning of 1968 to the end of 1983, the market price of gold rose from $35 to about $380 an ounce, which amounts to a 16.1 percent per annum nominal rate of increase. With an inflation rate of 7.1 percent over that time span, this means a 9.0 percent per annum "real" rate of return.

Over the same period, silver yielded a nominal 10.2 percent annual rate of return, stocks 8.1 percent, housing 8.0 percent, Treasury bills 7.5 percent, and long-term government bonds 5.3 percent.

In terms of wealth, $1 invested in gold at the start of 1968 would have grown to $10.84 by the end of 1983. Alternatively, $1 invested in silver would have grown to $4.74; $1 in stocks to $3.49; $1 in housing to $3.42; $1 in Treasury bills to $3.17; and $1 in long-term government bonds to $2.27.

With respect to riskiness, silver has been the most volatile and therefore most risky investment over the period, followed in order by gold, stocks, long-term government bonds, housing, and Treasury bills.

A generation ago, the stock market was widely viewed as a good hedge against inflation. However, since the mid-sixties it has yielded disappointing results. Between 1968 and 1983, on a year-to-year basis, inflation and stock prices moved in opposite directions more than twice as often as they moved in the same direction.

Gold, on the other hand, has done well as an inflation hedge. Based on monthly data over the period 1968-1983, the rate of return on gold and the rate of inflation have a correlation coefficient equal to .12; based on annual data over the same period, the correlation coefficient between gold and inflation is .52. In either case, the correlation between inflation and gold is higher than between inflation and any of the other assets considered in this study, with the exception of Treasury bills. Bills are the most reliable—that is, the least risky—inflation hedge of all, but they have a low real rate of return. Housing, long-term government bonds, and stocks have negative correlation coefficients with respect to inflation: when inflation went up, they went down.

Although gold is risky in itself—second only to silver among the assets considered—under many circumstances the addition of gold to a portfolio can *reduce* total portfolio risk. This is because the rate of return on gold has fluctuated opposite to the rate of return on stocks: when inflation accelerates, for example, the stock market typically falls while the price of gold typically rises.

For the period 1968-83, gold has a beta of .16 relative to the stock market as a whole, indicating that gold's returns have been less volatile than those of the stock market. The substitution of gold for part of an all-stock portfolio would thus have decreased total portfolio risk.

Typical consumers, however, do not hold an all-stock portfolio. They own their homes and often a fair amount of long-term bonds in their pension funds. For the 16 years 1968-83, gold has a beta of .41 calculated not with reference to the stock market as a whole but rather with reference to a total package consisting of one-third stocks, one-third long-term government bonds, and one-third housing. This bundle of assets is more representative of a typical consumer's investment portfolio than 100 percent investment in the stock market. The beta of .41 indicates that the substitution of gold for part of this portfolio would again have decreased total portfolio risk.

Thus although it is risky in and of itself, when added to a conventional portfolio gold can provide diversification that reduces rather than increases total portfolio risk. Since it has yielded a high rate of return, it can also enhance total earnings. Under the circumstances, it appears obvious that the historical record justifies a place for gold in a modern portfolio. How much of a place? How much gold should there be in a modern consumer portfolio?

We conclude that on average 5 to 10 percent of one's total investment portfolio in gold is warranted and prudent, although the proportion that is appropriate for any particular individual depends on such factors as how one views the future and how much risk one feels comfortable bearing.

THE MONOGRAPH SERIES IN FINANCE AND ECONOMICS

The following is a list of the available recent Monographs. The price is $5.00 per copy (except as otherwise noted).

1975— 1. Public and Private Offerings of Public Debt: Changes in the Yield Spread—Michel Fleuriet ($3.00)
2. Companies and Common Stocks as Inflation Hedges—Frank K. Reilly ($3.00)
3. "Managed Floating" As An Interim International Exchange rate Regime, 1973-1975—Samuel I. Katz ($3.00)
4. Improving the Access of Developing Countries to the U.S. Capital Market—Robert G. Hawkins, Walter L. Ness, Jr. and Il SaKong ($3.00)
5. Creating New Financial Instruments for the Aged—Jack M. Guttentag ($3.00)

1976— 1. The Stock Price Performance of Firms Repurchasing Their Own Shares—Kenneth R. Marks ($3.00)
2. The Liberalization of U.S. Capital Outflows: Internation Financial Consequences—Kaj Areskoug ($3.00)
3. Money, Credit, and Expenditure: A Sources and Uses of Funds Approach—James S. Earley. Robert T. Parsons, and Fred A. Thompson ($3.00)
4. The Political Economy of Postwar U.S. Trade Policy—Robert E. Baldwin ($3.00)

1977— 1. A Modest Proposal for a National Securities Market System and Its Governance—Ernest Bloch and Arnold W. Sametz ($3.00)
2. The Economic Effects of Compulsory Patent Licensing—F.M. Scherer ($3.00)
3. Profile of the New York Based Securities Industry—Michael Keenan ($3.00)
4. An Empirical Analysis of the Market for Tax-Exempt Securities: Estimates and Forecasts—Patric H. Hendershott and Timothy W. Koch ($3.00)

1978— 1. Tax Systems and Corporate Financing Policy—R.C. Stapleton and C.M. Burke (out of print)
2. Holding Period Yields and Risk-Primum Curves for Long-Term Marketable Securities: 1910-1976—Robert M. Soldofsky and Dale F. Max ($3.00)
3. The Behavior of Foreign Exchange Markets—A Critical Survey of the Empirical Literature—Steven W. Kohlhagen ($3.00)
4. Financial Effects of Capital Tax Reforms—Marshall E. Blume, Jean Crockett, and Irwin Friend($3.00)

5. An Empirical Study of Mortgage Redlining—George J. Benston, Dan Horsky and H. Martin Weingartner ($3.00)

1979— 1. The Implications of Managed Floating Exchange Rates for U.S. Trade Policy—Isaiah Frank, Charles Pearson and James Riedel ($3.00)

2. Regulation of Securities Markets: An Examination of the Effects of Increased Competition—Hans R. Stoll ($3.00)

3. Municipal Revenue Bond Costs and Bank Underwriting : A Survey of the Evidence—William L. Silber ($3.00)

4. Savings Bonds: Theory and Empirical Evidence—Michael J. Brennan and Eduardo S. Schwartz ($4.00)

1980— 1. Redlining and Public Policy—Jack M. Guttentag and Susan M. Wachter ($4.00)

2. Intraindustry Differences in Conduct and Performance: Viable Strategies in U.S. Manufacturing Industries—Richard E. Caves and Thomas A. Pugel ($4.00)

3. The Forward Market for Loanable Funds at Life Insurance Companies—Lawrence J. Horan ($4.00)

4. Discrimination in Mortgage Lending: A Study of Three Cities—A. Thomas King ($4.00)

1981— 1. Averting International Banking Crises—James W. Dean and Ian H. Giddy ($4.00)

2. The Regulation of Foreign Banks: Policy Formation In Countries Outside the United States—Adrian E. Tschoegl ($4.00)

3. The Analysis of Value and Volatility in Financial Futures—Martin Leibowitz ($4.00)

4. Measuring the Importance of Small Business in the American Economy—Lawrence J. White ($4.00)

1982— 1. Toward A Positive Theory of Insurance—David Mayers and Clifford W. Smith, Jr. ($4.00)

2. Do Stock Prices Reflect the Adjustment of Earnings for Inflation?—Phillip Cagan ($4.00)

3-4. Risk Arbitrage II—Guy P. Wyser-Pratte ($10)

1983— 1. Market, Firm and Economic Performance—Stephen Martin ($4.00)

2. Exchange-Rate Regime Selection In Theory and Practice—Gordon Weil ($4.00)

3. The Financial Impacts of Social Security—R. Glenn Hubbard ($4.00)

4. American Banks in the International Market—Stephen V. O. Clarke

1984— 1-2. The U.S. Payments System: Costs, Pricing, Competition and Risk—David Burras Humphrey ($7.50)

3. The Role of Gold in Consumer Investment Portfolios—Lawrence S. Ritter and Thomas J. Urich

SUBSCRIPTIONS

4 to 5 numbers issued per academic year-$15.00
(Special rate to full-time teachers-$7.50 per year)

Subscriptions and orders for individual Monographs should be sent to Salomon Brothers Center for the Study of Financial Institutions,
90 Trinity Place, New York, New York 10006

Out-of-print MONOGRAPHS and BULLETINS are available from:

University Microfilms International
Library Services
300 North Zeeb Road
Ann Arbor, Michigan 48106, U.S.A.

University Microfilms International
St. Johns Road, Tylers Green
High Wycombe
Buckinghamshire HP 10 8 HR, England

Inquiries should be made to the above addresses.

Customers in JAPAN may find it more convenient to order THE MONOGRAPH SERIES IN FINANCE AND ECONOMICS from the Far Eastern Book-Sellers, No. 2, 2-Chome, Kanda-Jinbocho, Chiyoda-Ku, Tokyo. This bookseller will usually have Salomon Brothers Center's Monographs in stock.

New York University
A private university in the public service

Graduate School of Business Administration

New York University is an affirmative action/equal opportunity institution.